GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 09:46:06 ; Search time 1142.55 Seconds
(without alignments)
9066.559 Million cell updates/sec

Title: US-10-614-282-1
Perfect score: 239
Sequence: 1 atcagtcacgacttggtaag.....tgtggttccggctggacaat 239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	239	100.0	1846	3	DMLABG1	X13104 Drosophila
2	239	100.0	2791	3	AY095089	AY095089 Drosophil
3	239	100.0	63442	2	AC020280	AC020280 Drosophil
c 4	239	100.0	85757	3	AC001652	AC001652 Drosophil
5	239	100.0	110000	3	AE001572_3	Continuation (4 of
c 6	239	100.0	170801	3	AC095014	AC095014 Drosophil
c 7	239	100.0	175648	3	AC095013	AC095013 Drosophil
c 8	239	100.0	298020	3	AE003674	AE003674 Drosophil
9	36.6	15.3	237630	2	AC106665	AC106665 Rattus no
10	36.6	15.3	239015	2	AC131194	AC131194 Rattus no
c 11	36.6	15.3	282019	2	AC096362	AC096362 Rattus no
12	34.8	14.6	20938	1	AE008718	AE008718 Salmonell
c 13	34	14.2	178993	2	AC099723	AC099723 Mus muscu
14	33.4	14.0	193636	2	AC113984	AC113984 Mus muscu
15	33.4	14.0	229102	2	AC119849	AC119849 Mus muscu
c 16	33.2	13.9	172214	2	BX294181	BX294181 Danio rer
17	33.2	13.9	240050	1	AL627267	AL627267 Salmonell
c 18	33.2	13.9	300029	1	AE016842	AE016842 Salmonell
c 19	32.8	13.7	727	6	AX416024	AX416024 Sequence
c 20	32.8	13.7	11621	1	AE003903	AE003903 Xylella f
c 21	32.8	13.7	151733	2	AC136249	AC136249 Rattus no
c 22	32.8	13.7	211426	2	AC119522	AC119522 Rattus no
c 23	32.8	13.7	216886	2	AC127846	AC127846 Rattus no
24	32.8	13.7	232601	2	AC112281	AC112281 Rattus no
c 25	32.8	13.7	232601	2	AC112281	AC112281 Rattus no
26	32.8	13.7	233002	2	AC120490	AC120490 Rattus no
27	32.8	13.7	234803	2	AC094061	AC094061 Rattus no
28	32.8	13.7	236723	2	AC106621	AC106621 Rattus no
c 29	32.8	13.7	244293	2	AC133774	AC133774 Rattus no
30	32.6	13.6	87077	5	BX571732	BX571732 Zebrafish
c 31	32.6	13.6	115859	8	CNS08CE0	BX000510 Oryza sat
32	32.6	13.6	129179	2	AC117205	AC117205 Mus muscu
33	32.6	13.6	134334	10	AL163512	AL163512 Mouse DNA
34	32.6	13.6	144486	10	AL626776	AL626776 Mouse DNA
c 35	32.6	13.6	148716	5	AL935049	AL935049 Zebrafish
36	32.6	13.6	165860	10	AL831716	AL831716 Mouse DNA
37	32.6	13.6	172465	2	BX324192	BX324192 Mus muscu
38	32.6	13.6	193230	2	AC138639	AC138639 Mus muscu
c 39	32.6	13.6	231548	2	AC146669	AC146669 Otolemur
40	32.6	13.6	246647	5	AL928716	AL928716 Zebrafish
c 41	32.4	13.6	100919	9	AL359378	AL359378 Human DNA
c 42	32.4	13.6	169226	2	AC107725	AC107725 Mus muscu
c 43	32.4	13.6	248818	2	AC112853	AC112853 Rattus no
c 44	32.4	13.6	256414	2	AC132775	AC132775 Rattus no
c 45	32.4	13.6	258710	2	AC135823	AC135823 Rattus no

ALIGNMENTS

RESULT 1
DMLABG1
LOCUS DMLABG1 1846 bp DNA linear INV 06-OCT-1999
DEFINITION Drosophila melanogaster lab gene for labial protein, exon 1 and joined CDS.
ACCESSION X13104 X12834
VERSION X13104.1 GI:8171
KEYWORDS alternative splicing; antennapedia complex; homeobox; lab gene.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1846)
AUTHORS Mlodzik,M., Fjose,A. and Gehring,W.J.

```
  TITLE     Molecular structure and spatial expression of a homeobox gene from
            the labial region of the Antennapedia-complex
  JOURNAL   EMBO J. 7 (8), 2569-2578 (1988)
  MEDLINE   89052673
   PUBMED   2461299
REFERENCE   2  (bases 1 to 1846)
  AUTHORS   Doelz,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1988) Doelz R., Biocomputing Biozentrum der
            Universiteit, Klingelbergstrasse 70, CH-4056 Basel, Switzerland
COMMENT     see M13570 for overlapping sequence.
FEATURES             Location/Qualifiers
     source          1. .1846
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="Oregon-R"
                     /db_xref="taxon:7227"
                     /chromosome="84A"
     gene            join(392. .1846,X13105.2:41. .474,X13106.1:1. .258)
                     /gene="lab"
     exon            392. .1846
                     /gene="lab"
                     /number=1
     CDS             join(631. .1846,X13105.2:41. .474,X13106.1:1. .258)
                     /gene="lab"
                     /codon_start=1
                     /label=lab_CDSI
                     /product="labial protein"
                     /protein_id="CAB57786.1"
                     /db_xref="GI:6015508"
                     /db_xref="GOA:P10105"
                     /db_xref="SWISS-PROT:P10105"
                     /translation="MMDVSSMYGNHPHHHHPHANAYDGYSTTTASAANASSYFAPQQH
                     QPHLQLQQQQQHQHLQQPQQHLTYNGYESSSPGNYYPQQQAQLTPPPTSSHQVVQQHQ
                     QQQQAQQQQLYPHSHLFSPSAAEYGITTSTTTGNPGKPLHPSSHSPADSYYESDSVHS
                     YYATAAVATVAPPSNSSPITAANASATSNTQQQQQQAAIISSENGMMYTNLDCMYPTA
                     QAQAPVHGYAGQIEEKYAAVLHASYAPGMVLEDQDPMMQQATQSQMWHHQQHLAGSYA
                     LDAMDSLGMHAHMHHGLPHGHLGNLANNPHQQQPQVQQQQQQPHQQPQHPQNQSPAAH
                     QQHHQNSVSPNGGMNRQQRGGVISPGSSTSSSTSASNGAHPASTQSKSPNHSSSIPTY
                     KWMQLKRNVPKPQAPSYLPAPKLPASGIASMHDYQMNGQLDMCRGGGGGGSDVGSGPV
                     GVGGNGSPGIGGVLSVQNSLIMANSAAAAGSAHPNGMGVGLGSGSGLSSCSLSSNTNN
                     SGRTNFTNKQLTELEKEFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRMKQKKRV
                     KEGLIPADILTQHSTSVISEKPPQQQQPQPPELQLKSQGSDLGGNELATGAPSTPTTA
                     MTLTAPTSKQS"
     CDS             join(631. .1846,X13105.2:59. .474,X13106.1:1. .258)
                     /gene="lab"
                     /codon_start=1
                     /label=lab_CDSII
                     /product="labial protein"
                     /protein_id="CAB57787.1"
                     /db_xref="GI:6015509"
                     /db_xref="GOA:P10105"
                     /db_xref="SWISS-PROT:P10105"
                     /translation="MMDVSSMYGNHPHHHHPHANAYDGYSTTTASAANASSYFAPQQH
                     QPHLQLQQQQQHQHLQQPQQHLTYNGYESSSPGNYYPQQQAQLTPPPTSSHQVVQQHQ
                     QQQQAQQQQLYPHSHLFSPSAAEYGITTSTTTGNPGKPLHPSSHSPADSYYESDSVHS
                     YYATAAVATVAPPSNSSPITAANASATSNTQQQQQQAAIISSENGMMYTNLDCMYPTA
                     QAQAPVHGYAGQIEEKYAAVLHASYAPGMVLEDQDPMMQQATQSQMWHHQQHLAGSYA
                     LDAMDSLGMHAHMHHGLPHGHLGNLANNPHQQQPQVQQQQQQPHQQPQHPQNQSPAAH
                     QQHHQNSVSPNGGMNRQQRGGVISPGSSTSSSTSASNGAHPASTQSKSPNHSSSIPTY
                     KWMQLKRNVPKPQAPKLPASGIASMHDYQMNGQLDMCRGGGGGGSDVGSGPVGVGGNG
                     SPGIGGVLSVQNSLIMANSAAAAGSAHPNGMGVGLGSGSGLSSCSLSSNTNNSGRTNF
                     TNKQLTELEKEFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRMKQKKRVKEGLIP
                     ADILTQHSTSVISEKPPQQQQPQPPELQLKSQGSDLGGNELATGAPSTPTTAMTLTAP
                     TSKQS"
ORIGIN

  Query Match           100.0%;  Score 239;  DB 3;  Length 1846;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-66;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        392 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 451

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        452 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 511

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        512 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 571

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        572 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 630


RESULT 2
AY095089
LOCUS       AY095089                2791 bp    mRNA    linear   INV 15-APR-2002
DEFINITION  Drosophila melanogaster RE63854 full insert cDNA.
ACCESSION   AY095089
VERSION     AY095089.1  GI:20152114
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2791)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
            Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
            and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2002) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone.  However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .2791
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /map="84A2-84A3"
     gene            1. .2791
                     /gene="lab"
                     /note="alignment with genomic scaffold AE003674"
                     /db_xref="FLYBASE:FBgn0002522"
     CDS             245. .2134
                     /gene="lab"
                     /note="Longest ORF"
                     /codon_start=1
                     /product="RE63854p"
```

```
                     /protein_id="AAM11417.1"
                     /db_xref="GI:20152115"
                     /db_xref="FLYBASE:FBgn0002522"
                     /translation="MMDVSSMYGNHPHHHHPHANAYDGYSTTTASAANASSYFAPQQH
                     QPHLQLQQQQQHQHLQQPQQHLTYNGYESSSPGNYYPQQQAQLTPPPTSSHQVVQQHQ
                     QQQQAQQQQLYPHSHLFSPSAAEYGITTSTTTGNPGTPLHPSSHSPADSYYESDSVHS
                     YYATAAVATVAPPSNSSPITAANASATSNTQQQQQQAAIISSENGMMYTNLDCMYPTA
                     QAQAPVHGYAGQIEEKYAAVLHASYAPGMVLEDQDPMMQQATQSQMWHHQQHLAGSYA
                     LDAMDSLGMHAHMHHGLPHGHLGNLANNPHQQQPQVQQQQQQPHQQPQHPQNQSPAAH
                     QQHHQNSVSPNGGMNRQQRGGVISPGSSTSSSTSASNGAHPASTQSKSPNHSSSIPTY
                     KWMQLKRNVPKPQAPKLPASGIASMHDYQMNGQLDMCRGGGGGGSGVGNGPVGVGGNG
                     SPGIGGVLSVQNSLIMANSAAAAGSAHPNGMGVGLGSGSGLSSCSLSSNTNNSGRTNF
                     TNKQLTELEKEFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRMKQKKRVKEGLIP
                     ADILTQHSTSVISEKPPQQQQPQPPELQLKSQGSDLGGNELATGAPSTPTTAMTLTAP
                     TSKQS"
ORIGIN

  Query Match           100.0%;  Score 239;  DB 3;  Length 2791;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-66;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 65

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 125

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 185

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 244


RESULT 3
AC020280
LOCUS       AC020280               63442 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC020280
VERSION     AC020280.1  GI:6664617
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 63442)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212650 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .63442
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

  Query Match           100.0%;  Score 239;  DB 2;  Length 63442;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-66;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6837 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 6896

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6897 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 6956

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6957 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 7016

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7017 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 7075


RESULT 4
AC001652/c
LOCUS       AC001652               85757 bp    DNA     linear   INV 17-JUL-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS00004 (D12)), complete
            sequence.
ACCESSION   AC001652 L76986 L32616 L46822 L81326 L79833 L32614 L81325 L42048
            L79834 L46820 L32619 L43416 L32618 L32615 L42053 L42049 L42051
            L42050 L76988 L42052 L76987 L81327 L32621 L46821 L32612 L43415
            L76985 L79835 L32622 L32613 L32617 L32620
VERSION     AC001652.1  GI:2341048
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 85757)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
  TITLE     Sequencing of antennapedia complex, homeotic genes
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 85757)
  AUTHORS   Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1997)
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library location: 4-1.
FEATURES             Location/Qualifiers
     source          1. .85757
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="84A1-84A2"
                     /clone="P1 DS00004 (D12)"
                     /note="This sequence has not changed since its original
                     submission on 04/22/1997.  It was resubmitted in order to
                     include all secondary accession numbers for the subclones
```

belonging to this clone."
ORIGIN

Query Match 100.0%; Score 239; DB 3; Length 85757;
Best Local Similarity 100.0%; Pred. No. 4.9e-66;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51586 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 51527

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51526 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 51467

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51466 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 51407

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51406 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 51348
```

RESULT 5
AE001572_3
WPCOMMENT
Sequence split into 5 fragments LOCUS AE001572 Accession AE001572

Fragment Name	Begin	End
AE001572_0	1	110000
AE001572_1	100001	210000
AE001572_2	200001	310000
AE001572_3	300001	410000
AE001572_4	400001	429825

Continuation (4 of 5) of AE001572 from base 300001 (AE001572 Drosophila melanogaster Ant

Query Match 100.0%; Score 239; DB 3; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.1e-66;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78240 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 78299

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78300 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 78359

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78360 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 78419

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78420 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 78478
```

RESULT 6
AC095014/c
LOCUS AC095014 170801 bp DNA linear INV 15-SEP-2001
DEFINITION Drosophila melanogaster, chromosome 3R, region 83F-84A, BAC clone BACR19D23, complete sequence.
ACCESSION AC095014
VERSION AC095014.1 GI:15624856
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 170801)
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE Sequencing of Drosophila chromosome 3R, region 83F-84A
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170801)
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US
COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu.
FEATURES Location/Qualifiers
source 1. .170801
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="83F-84A"
/clone="BACR19D23 (D1349)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACe3.6)"
ORIGIN

Query Match 100.0%; Score 239; DB 3; Length 170801;
Best Local Similarity 100.0%; Pred. No. 5.5e-66;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12530 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 12471

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12470 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 12411

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12410 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 12351

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      12350 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 12292


RESULT 7
AC095013/c
LOCUS       AC095013              175648 bp    DNA     linear   INV 15-SEP-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 83F-83F, BAC clone
            BACR19I14, complete sequence.
ACCESSION   AC095013
VERSION     AC095013.1  GI:15624855
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 175648)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 83F-83F
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175648)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .175648
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="83F-83F"
                     /clone="BACR19I14 (D1348)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
ORIGIN
```

```
  Query Match           100.0%;  Score 239;  DB 3;  Length 175648;
  Best Local Similarity  100.0%;  Pred. No. 5.5e-66;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144223 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 144164

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144163 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 144104

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144103 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 144044

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144043 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 143985


RESULT 8
AE003674/c
LOCUS       AE003674              298020 bp    DNA     linear   INV 14-FEB-2003
DEFINITION  Drosophila melanogaster chromosome 3R, section 10 of 118 of the
            complete sequence.
ACCESSION   AE003674 AE002699 AE014297
VERSION     AE003674.3  GI:23170630
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 298020)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,
            Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,
            Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,
            Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S.,
            Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P.,
            Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A.,
            Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
            Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
            Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,
            Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
            Ferriera,S., Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., WoodageT,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
```

```
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
   PUBMED   10731132
REFERENCE   2  (bases 1 to 298020)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Wan,K.H., Holt,R.A.,
            Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y.,
            Banzon,J., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Tyler,D.,
            Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster genome
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 298020)
  AUTHORS   Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,
            Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochnik,S.E.,
            Smith,C.D., Tupy,J.L., Bergman,C.M., Berman,B.P., Carlson,J.W.,
            Celniker,S.E., Clamp,M.E., Drysdale,R.A., Emmert,D., Frise,E., de
            Grey,A.D.N.J., Harris,N.L., Kronmiller,B., Marshall,B.,
            Millburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E.,
            Shu,S., Smutniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M.,
            Rubin,G.M., Mungall,C.J. and Lewis,S.E.
  TITLE     Annotation of Drosophila melanogaster genome
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 298020)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
REFERENCE   5  (bases 1 to 298020)
  AUTHORS   .
  CONSRTM   FlyBase
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2002) University of California Berkeley, 539 Life
            Sciences Addition, Berkeley, CA 94720, USA
REFERENCE   6  (bases 1 to 298020)
  AUTHORS   .
  CONSRTM   FlyBase
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2003) University of California Berkeley, 539 Life
            Sciences Addition, Berkeley, CA 94720, USA
  REMARK    Sequence update by submitter
COMMENT     On Sep 18, 2002 this sequence version replaced gi:10727121.
FEATURES             Location/Qualifiers
     source          1. .298020
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /note="genotype: y[1]; cn[1] bw[1] sp[1]; Rh6[1]"
     gene            2578. .14815
                     /locus_tag="CG15186"
                     /note="last curated on Thu Jan 03 10:40:34 PST 2002"
                     /map="83E8-83F1"
                     /db_xref="FLYBASE:FBgn0037448"
     mRNA            join(2578. .2960,9409. .9497,9940. .10163,10305. .10755,
                     10823. .11763,12392. .13156,13329. .14815)
                     /locus_tag="CG15186"
                     /product="CG15186-RA"
                     /db_xref="FLYBASE:FBgn0037448"
     mRNA            join(6771. .7105,9409. .9497,9940. .10163,10305. .10755,
                     10823. .11763,12392. .13156,13329. .14815)
                     /locus_tag="CG15186"
                     /product="CG15186-RB"
                     /db_xref="FLYBASE:FBgn0037448"
     CDS             join(9434. .9497,9940. .10163,10305. .10755,10823. .11763,
                     12392. .13156,13329. .14054)
                     /locus_tag="CG15186"
                     /note="CG15186 gene product from transcript CG15186-RA"
                     /codon_start=1
                     /product="CG15186-PA"
                     /protein_id="AAF54118.1"
                     /db_xref="GI:7298912"
                     /db_xref="FLYBASE:FBgn0037448"
                     /translation="MIYENEELRLRTININAEVERGQSDIKRLRRENEQLRREIWTLR
                     DECDRLNKRFKAKLNEHEFGGCRGSGGSGGTCHAYRCSGGGRGSGGCDANSDDSDSCD
                     TCRGADDGHCSDECCQEGGSCGPPKPPLPPEVPSHSSQSKNEDGNSSNQTKEPQPMHF
                     DHLSVVSEETLSNPELMHVQQQEHLTIFPPTDAHGSQSTLPSMMGPLTPLTPIELVAN
                     QLNDLQAMVPPLSYFENILSQHMGARNAAQTPSPELGTSSSTTTGKTMRHTNGWDYNL
                     QSPFTQRKYNDPSSPSRSPPPPVSTMTTFVPTSTLQTMPKIALNAPPQPHLADGGDIQ
                     TVAITTNAAQHALNSPKHFFAPIKPRLKLNTKLANQGQDLEQDDLPPGSPPPPLRDPP
                     DIFVNNALASPYQRQVSPQVSPRHRRSTHSHCHARSHQRSRQRPQPSHCSLHRAVVDV
                     AAAVGLGEVGPHSGSSVLDIYTAALAAAAAVRCSPLPAARSVEPIYATAQKPCQNMCI
                     SATSVTQPVMTITTQSNKTTTTRWPSPKRLASRAKAVGEAKGAAKSAENVSQTDSVNL
                     ESILNDIETISEDILAIQLEKSKSRDNLSHNNNKDDERAKKPYRSEMNLYLQYDGTNP
                     TISPATTDTQTTEVGTTTPAVTTSSESGNSGSNKLVRRTRSLEREHTDSPAPHIPDPM
                     APFPDKCTYLGFEQLNQAAAGSGVPPQSPNDQRPPIAAKPNVPLKPPPPLPPARRPTM
                     PTEPPPSPPPTTSASGVIGLPPNKSHLYKTLASAAAKRAIFRSSPSQLTRSLDVDPAV
                     PDGEVGVAGDQSSAAANMDELTRRKARRVSIVCGSGQGGQQQDDHVTTPVTTVPMATSS
                     CVDLSTVLTHPAIKAFKMSHSTPNSPHSSRRRTNSNSMSAPPGIPVSSGHPEVGATAS
                     APPSTTHHHHHRKEGSDPVPMTATVSRTKCSRRHSEGTVHTVHRNSAASGGAGGGHHH
                     HSHSHHHPHSGMVISSNPHHSHHSHQDSNHETVTSLSDRNSNSFASSRESSTSFSMRS
                     NRRKISVSSHTGGKIPWCGCWGNGCL"
     CDS             join(9434. .9497,9940. .10163,10305. .10755,10823. .11763,
                     12392. .13156,13329. .14054)
                     /locus_tag="CG15186"
                     /note="CG15186 gene product from transcript CG15186-RB"
                     /codon_start=1
                     /product="CG15186-PB"
                     /protein_id="AAN13360.1"
                     /db_xref="GI:23170631"
                     /db_xref="FLYBASE:FBgn0037448"
                     /translation="MIYENEELRLRTININAEVERGQSDIKRLRRENEQLRREIWTLR
                     DECDRLNKRFKAKLNEHEFGGCRGSGGSGGTCHAYRCSGGGRGSGGCDANSDDSDSCD
                     TCRGADDGHCSDECCQEGGSCGPPKPPLPPEVPSHSSQSKNEDGNSSNQTKEPQPMHF
                     DHLSVVSEETLSNPELMHVQQQEHLTIFPPTDAHGSQSTLPSMMGPLTPLTPIELVAN
                     QLNDLQAMVPPLSYFENILSQHMGARNAAQTPSPELGTSSSTTTGKTMRHTNGWDYNL
                     QSPFTQRKYNDPSSPSRSPPPPVSTMTTFVPTSTLQTMPKIALNAPPQPHLADGGDIQ
                     TVAITTNAAQHALNSPKHFFAPIKPRLKLNTKLANQGQDLEQDDLPPGSPPPPLRDPP
                     DIFVNNALASPYQRQVSPQVSPRHRRSTHSHCHARSHQRSRQRPQPSHCSLHRAVVDV
                     AAAVGLGEVGPHSGSSVLDIYTAALAAAAAVRCSPLPAARSVEPIYATAQKPCQNMCI
                     SATSVTQPVMTITTQSNKTTTTRWPSPKRLASRAKAVGEAKGAAKSAENVSQTDSVNL
                     ESILNDIETISEDILAIQLEKSKSRDNLSHNNNKDDERAKKPYRSEMNLYLQYDGTNP
                     TISPATTDTQTTEVGTTTPAVTTSSESGNSGSNKLVRRTRSLEREHTDSPAPHIPDPM
                     APFPDKCTYLGFEQLNQAAAGSGVPPQSPNDQRPPIAAKPNVPLKPPPPLPPARRPTM
                     PTEPPPSPPPTTSASGVIGLPPNKSHLYKTLASAAAKRAIFRSSPSQLTRSLDVDPAV
                     PDGEVGVAGDQSSAAANMDELTRRKARRVSIVCGSGQGGQQQDDHVTTPVTTVPMATSS
                     CVDLSTVLTHPAIKAFKMSHSTPNSPHSSRRRTNSNSMSAPPGIPVSSGHPEVGATAS
                     APPSTTHHHHHRKEGSDPVPMTATVSRTKCSRRHSEGTVHTVHRNSAASGGAGGGHHH
                     HSHSHHHPHSGMVISSNPHHSHHSHQDSNHETVTSLSDRNSNSFASSRESSTSFSMRS
                     NRRKISVSSHTGGKIPWCGCWGNGCL"
     gene            complement(15821. .15893)
                     /gene="tRNA:M2:83F"
                     /locus_tag="CR31310"
                     /map="83F1-83F1"
     tRNA            complement(15821. .15893)
                     /gene="tRNA:M2:83F"
                     /locus_tag="CR31310"
                     /product="tRNA-Met"
                     /note="last curated on Fri Mar 22 14:06:17 PST 2002"
                     /db_xref="FLYBASE:FBgn0011919"
     gene            29863. .34879
                     /locus_tag="TE19320"
                     /map="83F1-83F1"
     repeat_region   29863. .34879
                     /transposon="jockey{}1238"
                     /db_xref="FLYBASE:FBti0019320"
     gene            47311. .48339
                     /locus_tag="CG15185"
```

```
                     /note="last curated on Thu Jan 03 10:58:20 PST 2002"
                     /map="83F1-83F1"
                     /db_xref="FLYBASE:FBgn0037449"
     mRNA            47311. .48339
                     /locus_tag="CG15185"

  Query Match             100.0%;  Score 239;  DB 3;  Length 298020;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-66;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198367 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 198308

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198307 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 198248

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198247 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 198188

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198187 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 198129


RESULT 9
AC106665
LOCUS       AC106665               237630 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-102I13, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC106665
VERSION     AC106665.4  GI:30578557
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 237630)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 237630)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 237630)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:22857405.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLTZ
                Center clone name: CH230-102I13
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 229128 bases at least Q40
                Consensus quality: 231415 bases at least Q30
                Consensus quality: 233051 bases at least Q20
                Estimated insert size: 240583; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   234954: contig of 234954 bp in length
            *  234955   235054: gap of unknown length
            *  235055   236441: contig of 1387 bp in length
```

```
           *   236442   236541: gap of unknown length
           *   236542   237630: contig of 1089 bp in length.
FEATURES             Location/Qualifiers
     source          1. .237630
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-102I13"
     misc_feature    1. .1505
                     /note="wgs_contig"
     misc_feature    2205. .4409
                     /note="wgs_contig"
ORIGIN

  Query Match             15.3%;  Score 36.6;  DB 2;  Length 237630;
  Best Local Similarity   58.9%;  Pred. No. 4.1;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        120 ATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGG 179
              ||| | || |||| ||| || ||| | || ||||| |  ||| | |||| ||    ||
Db     215052 ATAGTTTTGGAAACCAAGCAACTAGATAAATATTTTATCAACAGAGTGGAAATGCTAGAA 215111

Qy        180 TAGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTT 226
              |||    | | |   | ||  | || || || | || |  |   | ||
Db     215112 TAGAGCATTAGATGGCTGTGTTTTCTGAATGACAAGGAGTTCAGCTT 215158


RESULT 10
AC131194
LOCUS       AC131194               239015 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-85M5, WORKING DRAFT SEQUENCE.
ACCESSION   AC131194
VERSION     AC131194.3  GI:30522034
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 239015)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 239015)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 239015)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23264076.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGBH
                Center clone name: CH230-85M5
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 224202 bases at least Q40
                Consensus quality: 226474 bases at least Q30
                Consensus quality: 228026 bases at least Q20
                Estimated insert size: 235880; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   239015: contig of 239015 bp in length.
FEATURES             Location/Qualifiers
     source          1. .239015
```

```
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-85M5"
     misc_feature    1. .1575
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    10013. .10837
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH305681"
     misc_feature    complement(235727. .236614)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH305680"
ORIGIN

  Query Match             15.3%;  Score 36.6;  DB 2;  Length 239015;
  Best Local Similarity   58.9%;  Pred. No. 4.1;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        120 ATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGG 179
              ||| | || |||| ||| || ||| | || ||||| |  ||| | |||| ||    ||
Db      14324 ATAGTTTTGGAAACCAAGCAACTAGATAAATATTTTATCAACAGAGTGGAAATGCTAGAA 14383

Qy        180 TAGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTT 226
              |||    | | |   | ||  | || || || | || |   |  | ||
Db      14384 TAGAGCATTAGATGGCTGTGTTTTCTGAATGACAAGGAGTTCAGCTT 14430


RESULT 11
AC096362/c
LOCUS       AC096362              282019 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-53N24, WORKING DRAFT SEQUENCE.
ACCESSION   AC096362
VERSION     AC096362.6  GI:30521829
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 282019)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 282019)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 282019)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:22771585.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEZO
                Center clone name: CH230-53N24
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 236414 bases at least Q40
                Consensus quality: 239934 bases at least Q30
                Consensus quality: 242193 bases at least Q20
                Estimated insert size: 249348; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
```

```
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1   282019: contig of 282019 bp in length.
FEATURES             Location/Qualifiers
     source          1. .282019
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-53N24"
     misc_feature    4444. .5560
                     /note="wgs_contig"
     misc_feature    279586. .280897
                     /note="wgs_contig"
     misc_feature    280948. .282019
                     /note="wgs_contig"
ORIGIN

  Query Match             15.3%;  Score 36.6;  DB 2;  Length 282019;
  Best Local Similarity   58.9%;  Pred. No. 4.2;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        120 ATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGG 179
              ||| | || |||| ||| || ||| | || |||||  |  ||| | |||| ||    ||
Db     258385 ATAGTTTTGGAAACCAAGCAACTAGATAAATATTTTATCAACAGAGTGGAAATGCTAGAA 258326

Qy        180 TAGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTT 226
              |||    | | |   | ||  | || || || | || |  |  | ||
Db     258325 TAGAGCATTAGATGGCTGTGTTTTCTGAATGACAAGGAGTTCAGCTT 258279


RESULT 12
AE008718
LOCUS       AE008718               20938 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 26 of 220 of the complete
            genome.
ACCESSION   AE008718 AE006468
VERSION     AE008718.1  GI:16418985
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 20938)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 20938)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .20938
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            90. .3452
                     /gene="aefA"
                     /note="synonym: STM0478"
     CDS             90. .3452
                     /gene="aefA"
                     /note="similar to E. coli putative alpha helix protein
                     (AAC73567.1); Blastp hit to AAC73567.1 (1120 aa), 89%
                     identity in aa 1 - 1117"
                     /codon_start=1
                     /transl_table=11
                     /product="putative small-conductance mechanosensitive
                     channel"
                     /protein_id="AAL19432.1"
                     /db_xref="GI:16418986"
                     /translation="MTMLQLYKRSQHFVFITISVLIILLSCQSLAFARGQTNGDLPSK
                     ADVQNQLDTLNKQKDLSAQDKLVQQDLIDTLATLEKIERVKEETVQLRQKVAQAPEKM
                     RQATAALNALSDVDNDDEMRKTLSALSLRQLELRVAQVLDDLQNSQNDLAAYNSQLVS
                     LQTQPERVQNAMYTASQQIQQIRNRLDGNNVGEAALRPSQQVLLQAQQALLNAQIDQQ
                     RKSLEGNTVLQDTLQKQRDYVTANSNRLEHQLQLLQEAVNSKRLTLTEKTAQEAISPD
                     ETARIQANPLVKQELDINHQLSQRLIVATENGNMLMQQNIKVKNWLDRALQSERNIKE
                     QIAVLKGSLLLSRILYQQQQTLPSADELEDMTNRIADLRLEQFEINQQRDALFQSDAF
                     VDKLEEGHTSEVNDEVHDALLQVVEMRRELLDQLNKQLGNQLMMAINLQVNQQQLMSV
                     SKNLKAILTQQIFWVNSNRPMDWDWLKAFPQTLKEQFSAMKITVNWQKAWPAVFIAFL
                     AGLPLLLIAGLIRWRLKWLKAYQQKLAAAVGSLRNDSQLNTPKAILIDLIRALPVCLI
                     ILALGLILLTMQLNISDLLWAFSKKLAMFWLVFGLCWKVLEKEGVAIRHFGMPAQLTS
                     HWRRQIVRISLALLPLHFWSVVAELSPLNLMDDVLGQAVIFLNLLVITLLVWPLCRES
                     WRDKESHGIRLVTVTILSIIPVALMVLTATGYFYTTLRLAGRWIETVYLVIIWNLLYQ
                     TVLRGLSVAARRIAWRRALARRQNLVKEGAEGAEPQEEPTIALEQINQQTLRITMLLM
                     LALFGVMFWAIWSDLITVFSYLDSITLWHYNGSEAGAAVVKSVTMGSLLFAIIAAMVA
                     WALIRNLPGLLEVLVLSRLNMRQGASYAITTILNYVIIAVGAMTVFGSLGVSWDKLQW
                     LAAALSVGLGFGLQEIFGNFVSGLIILFERPVRIGDTVTIGTYSGTVSKIRIRATTIT
                     DFDRKEVIIPNKAFVTERLINWSLSDTTTRLVIRLGVAYGSDLEKVKRVLLQAAMEHP
                     KVMHDPEPAVFFTTFGASTLDHELRLYVRELRDRSHTVDELNRAIDRLCRENDINIAF
                     NQLEVHLHNAKGDEVTEVKRDLNGGDLAPTAS"
     gene            complement(3494. .4435)
                     /gene="STM0479"
     CDS             complement(3494. .4429)
                     /gene="STM0479"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC75365.1); Blastp hit to AAC75365.1 (296 aa), 49%
                     identity in aa 7 - 296"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAL19433.1"
                     /db_xref="GI:16418987"
                     /translation="MERLPTTPHDAVFRQMLMQKEVARDFLAIHMPEDFLAICDLDSL
                     KLESGSFVEDNLRSRYSDILYSLHTQHGPGYVYALIEHQSKSDRLMAFRLMRYAIAAM
                     QRHLDAGHDTLPLVVPILFYHGPESPWPYSLNWHNMFVKPDMAKALYSRDFALVDLTT
                     MPDNQLLQHRRIAMLELLQKHIRQRDLSELLDPLITLLTQDHLTDAQLSVLINYMLKA
                     GNAAEPGALIRQLAQGAPQYKEQLMTIAEWLEEKGRTEGLQKGLQKGLEQGLAQGREA
                     EARAIARKMLANGLEPGLIASVTGITPEELSTLSH"
     RBS             complement(4430. .4435)
                     /gene="STM0479"
                     /note="putative RBS for STM0479; RegulonDB:STMS1H000821"
```

```
gene            complement(4499. .4677)
                /gene="ybaM"
                /note="synonym: STM0480"
CDS             complement(4499. .4666)
                /gene="ybaM"
                /note="similar to E. coli orf, hypothetical protein
                (AAC73568.1); Blastp hit to AAC73568.1 (53 aa), 80%
                identity in aa 1 - 49"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL19434.1"
                /db_xref="GI:16418988"
                /translation="MSLENAPDEVKLAVDLIVLLEENRLPARTVLRALEIVMRDYENK
                LKSTEDDSQTE"
RBS             complement(4672. .4677)
                /gene="ybaM"
                /note="putative RBS for ybaM; RegulonDB:STMS1H000822"
gene            complement(4680. .5195)
                /gene="priC"
                /note="synonym: STM0481"
CDS             complement(4680. .5195)
                /gene="priC"
                /note="similar to E. coli primosomal replication protein
                N'' (AAC73569.1); Blastp hit to AAC73569.1 (175 aa), 73%
                identity in aa 5 - 175"
                /codon_start=1
                /transl_table=11
                /product="primosomal replication protein N"
                /protein_id="AAL19435.1"
                /db_xref="GI:16418989"
                /translation="MLLQTLEERLATLRQRCAPLAQHATLSARFDRHLFRTRSTLLQG
                YLEEAGANLVALRQAVKHEQLPQVAWLAEHLASQLEAISRETAAWSLRQWDAAAPGLG
                RWQRRRIQHQEFERRLLAMTQERKIRLAQATGLVEQQTLQKEVEIYEGRLARCRHALE
                KIENVLARLTR"
gene            5276. .5653
                /gene="ybaN"
                /note="synonym: STM0482"
CDS             5276. .5653
                /gene="ybaN"
                /note="similar to E. coli putative gene 58 (AAC73570.1);
                Blastp hit to AAC73570.1 (125 aa), 84% identity in aa 1 -
                124"
                /codon_start=1
                /transl_table=11
                /product="putative phage gene 58"
                /protein_id="AAL19436.1"
                /db_xref="GI:16418990"
                /translation="MQRTILIIIGWLAVVLGTLGVVLPLLPTTPFILLAAWCFARSSP
                RFHAWLLYRSWFGGYLRHWQRYRAMPPGAKPRAIALILLTFGISLWLVNMMWVRVLLL
                VILACLLIFMWRIPVIDEKQQKR"
gene            5793. .6357
                /gene="apt"
                /note="synonym: STM0483"
RBS             5793. .5798
                /gene="apt"
                /note="putative RBS for apt; RegulonDB:STMS1H000823"
CDS             5806. .6357
                /gene="apt"
                /EC_number="2.4.2.7"
                /note="similar to E. coli adenine
                phosphoribosyltransferase (AAC73571.1); Blastp hit to
                AAC73571.1 (183 aa), 95% identity in aa 1 - 183"
                /codon_start=1
                /transl_table=11
                /product="adenine phosphoribosyltransferase"
                /protein_id="AAL19437.1"
                /db_xref="GI:16418991"
                /translation="MTATAQQLEFLKNSIKSIQDYPKPGILFRDVTSLLEDPKAYALS
                IELLVERYKNAGITKVVGTEARGFLFGAPVALGLGVGFVPVRKPRKLPRETIAETYEL
                EYGTDQLEIHVDAIKPGDNVLVVDDLLATGGTIEATVKLIRRLGGKVTDAAFIINLFD
                LGGEQRLEKQGITCYSLVPFPGH"
gene            6471. .8399
                /gene="dnaX"
                /note="synonym: STM0484"
CDS             6471. .8399
                /gene="dnaX"
                /EC_number="2.7.7.7"
                /note="DNA elongation factor III; DNA polymerase III
                subunit TAU  [contains: DNA polymerase III subunit gamma].
                 (SW:DP3X_SALTY)"
                /codon_start=1
                /transl_table=11
                /product="DNA polymerase III, tau and gamma subunits"
                /protein_id="AAL19438.1"
                /db_xref="GI:16418992"

  Query Match            14.6%;  Score 34.8;  DB 1;  Length 20938;
  Best Local Similarity  52.0%;  Pred. No. 10;
  Matches   78;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         23 CGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACA 82
              |||||  ||||   ||  ||  |   |  ||  |    | |     |  | |  ||  |
Db      18666 CGCAGCGAGCAATACGCCGCCAAAACTTCCCGCCATTGCCAGCAGTACCAGCCCGCTCAG 18725

Qy         83 GATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCT 142
              | |||||| || | ||| || |  |||||   || |   || ||||     | |  |||
Db      18726 GGTAAAACACATGCCGACAGGCAGCGACCACGGCATGGGATACTTATCCGTCCAGTAGCC 18785

Qy        143 AAAGAACTATTTCAAGAACTGTGTGGCAAG 172
               | ||  |  || ||  | |  || ||| ||
Db      18786 GACGACCGGTTGCAGCAGCGATGAGGCGAG 18815


RESULT 13
AC099723/c
LOCUS       AC099723              178993 bp    DNA     linear   HTG 19-OCT-2002
DEFINITION  Mus musculus clone RP23-369P21, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC099723
VERSION     AC099723.2  GI:24137577
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 178993)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-369P21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178993)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```

```
REFERENCE   3  (bases 1 to 178993)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 19, 2002 this sequence version replaced gi:16974222.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18019
                Center clone name: 369_P_21
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 176482 bases at least Q40
                Consensus quality: 177487 bases at least Q30
                Consensus quality: 177924 bases at least Q20
                Insert size: 183000; agarose-fp
                Insert size: 178293; sum-of-contigs
                Quality coverage: 10.8 in Q20 bases; agarose-fp
                Quality coverage: 11.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8673: contig of 8673 bp in length
            *     8674     8773: gap of 100 bp
            *     8774     9858: contig of 1085 bp in length
            *     9859     9958: gap of 100 bp
            *     9959    19220: contig of 9262 bp in length
            *    19221    19320: gap of 100 bp
            *    19321    29525: contig of 10205 bp in length
            *    29526    29625: gap of 100 bp
            *    29626    95507: contig of 65882 bp in length
            *    95508    95607: gap of 100 bp
            *    95608   128597: contig of 32990 bp in length
            *   128598   128697: gap of 100 bp
            *   128698   172325: contig of 43628 bp in length
            *   172326   172425: gap of 100 bp
            *   172426   178993: contig of 6568 bp in length.
FEATURES            Location/Qualifiers
     source         1. .178993
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /clone="RP23-369P21"
                    /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1. .8673
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   8774. .9858
                    /note="assembly_fragment"
     misc_feature   9959. .19220
                    /note="assembly_fragment"
     misc_feature   19321. .29525
                    /note="assembly_fragment"
     misc_feature   29626. .95507
                    /note="assembly_fragment"
     misc_feature   95608. .128597
                    /note="assembly_fragment"
     misc_feature   128698. .172325
                    /note="assembly_fragment"
     misc_feature   172426. .178993
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
ORIGIN

  Query Match             14.2%;  Score 34;  DB 2;  Length 178993;
  Best Local Similarity   51.3%;  Pred. No. 27;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         83 GATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCT 142
              | | || | ||    | ||  ||| | || |  | | ||  ||| ||| || ||   |
Db     115682 GCTTAACCACAGAGCCATCTCCCCAGGCCTCCTGGAAGTAAACTTTTAAACCACACTTTT 115623

Qy        143 AAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGATACACCGGTTATA 202
              | ||  || ||| || |  ||| |  ||| |    |||    ||   | ||   ||  ||
Db     115622 ATAGTTCTTTTTAAAAATATGTATTTCAATTATGTGGTGAGGAGGACTGCAAATGTCCTA 115563

Qy        203 TCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGAC 236
                |  | ||| | |||       ||||| |
Db     115562 GAAGCTAGAAGAGGATGTGACCCCGGGGCTGGTC 115529


RESULT 14
AC113984
LOCUS       AC113984              193636 bp    DNA     linear   HTG 21-OCT-2002
DEFINITION  Mus musculus clone RP24-97D3, WORKING DRAFT SEQUENCE, 13 ordered
            pieces.
ACCESSION   AC113984
VERSION     AC113984.3  GI:24182209
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 193636)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-97D3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193636)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
```

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (06-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 193636)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Oct 21, 2002 this sequence version replaced gi:21327615.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L24343
Center clone name: 97_D_3
-------------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 189906 bases at least Q40
Consensus quality: 191355 bases at least Q30
Consensus quality: 191926 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 192436; sum-of-contigs
Quality coverage: 9.5 in Q20 bases; agarose-fp
Quality coverage: 8.3 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 1188: contig of 1188 bp in length
* 1189 1288: gap of 100 bp
* 1289 3601: contig of 2313 bp in length
* 3602 3701: gap of 100 bp
* 3702 6708: contig of 3007 bp in length
* 6709 6808: gap of 100 bp
* 6809 8666: contig of 1858 bp in length
* 8667 8766: gap of 100 bp
* 8767 13690: contig of 4924 bp in length
* 13691 13790: gap of 100 bp
* 13791 62565: contig of 48775 bp in length
* 62566 62665: gap of 100 bp
* 62666 66508: contig of 3843 bp in length
* 66509 66608: gap of 100 bp
* 66609 74500: contig of 7892 bp in length
* 74501 74600: gap of 100 bp
* 74601 97683: contig of 23083 bp in length
* 97684 97783: gap of 100 bp
* 97784 134290: contig of 36507 bp in length
* 134291 134390: gap of 100 bp
* 134391 158667: contig of 24277 bp in length
* 158668 158767: gap of 100 bp
* 158768 189834: contig of 31067 bp in length
* 189835 189934: gap of 100 bp
* 189935 193636: contig of 3702 bp in length.

FEATURES Location/Qualifiers

source 1. .193636
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-97D3"
/clone_lib="RPCI-24 Male Mouse BAC"

misc_feature 1. .1188
/note="assembly_fragment"

misc_feature 1289. .3601
/note="assembly_fragment"

misc_feature 3702. .6708
/note="assembly_fragment"

misc_feature 6809. .8666
/note="assembly_fragment"

misc_feature 8767. .13690
/note="assembly_fragment"

misc_feature 13791. .62565
/note="assembly_fragment"

misc_feature 62666. .66508
/note="assembly_fragment"

misc_feature 66609. .74500
/note="assembly_fragment"

misc_feature 74601. .97683
/note="assembly_fragment"

misc_feature 97784. .134290
/note="assembly_fragment"

misc_feature 134391. .158667
/note="assembly_fragment"

misc_feature 158768. .189834
/note="assembly_fragment"

misc_feature 189935. .193636
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN

Query Match 14.0%; Score 33.4; DB 2; Length 193636;
Best Local Similarity 54.5%; Pred. No. 43;
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy        50 GCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGA 109
             | | | ||| |  |||  |||| || | | |||| |  |  ||  | | | |||| |
Db      3256 GGCCAGGGGGGCTGTGGATTTCCTTTAGTTTCAGAGCACTCAGACTGCCAAATCCCTGTA 3315

Qy       110 CCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGC 169
             |  |    | |||| |  |    |||| |  ||| ||  |    |||  ||| |||||
Db      3316 CTCTGAGTTGTTATTTTCTGGGCCCAAAAAAAAAAAAAAAAAAAAAAAGGGCTGAGTGGC 3375

Qy       170 AAG 172
             |||
Db      3376 CAG 3378
```

RESULT 15

```
AC119849
LOCUS       AC119849               229102 bp    DNA     linear   HTG 22-FEB-2003
DEFINITION  Mus musculus clone RP23-74O4, WORKING DRAFT SEQUENCE, 12 ordered
            pieces.
ACCESSION   AC119849
VERSION     AC119849.2  GI:28467231
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 229102)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-74O4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229102)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 229102)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 22, 2003 this sequence version replaced gi:20389488.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22358
                Center clone name: 74_O_4
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 226842 bases at least Q40
                Consensus quality: 227526 bases at least Q30
                Consensus quality: 227772 bases at least Q20
                Insert size: 212000; agarose-fp
                Insert size: 228002; sum-of-contigs
                Quality coverage: 11.5 in Q20 bases; agarose-fp
                Quality coverage: 10.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    12780: contig of 12780 bp in length
            *    12781    12880: gap of 100 bp
            *    12881    13370: contig of 490 bp in length
            *    13371    13470: gap of 100 bp
            *    13471    14758: contig of 1288 bp in length
            *    14759    14858: gap of 100 bp
            *    14859    17963: contig of 3105 bp in length
            *    17964    18063: gap of 100 bp
            *    18064    20876: contig of 2813 bp in length
            *    20877    20976: gap of 100 bp
            *    20977    27426: contig of 6450 bp in length
            *    27427    27526: gap of 100 bp
            *    27527    41619: contig of 14093 bp in length
            *    41620    41719: gap of 100 bp
            *    41720    98140: contig of 56421 bp in length
            *    98141    98240: gap of 100 bp
            *    98241   120308: contig of 22068 bp in length
            *   120309   120408: gap of 100 bp
            *   120409   160843: contig of 40435 bp in length
            *   160844   160943: gap of 100 bp
            *   160944   216116: contig of 55173 bp in length
            *   216117   216216: gap of 100 bp
            *   216217   229102: contig of 12886 bp in length.
FEATURES             Location/Qualifiers
     source          1. .229102
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-74O4"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .12780
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    12881. .13370
                     /note="assembly_fragment"
     misc_feature    13471. .14758
                     /note="assembly_fragment"
     misc_feature    14859. .17963
                     /note="assembly_fragment"
     misc_feature    18064. .20876
                     /note="assembly_fragment"
     misc_feature    20977. .27426
                     /note="assembly_fragment"
     misc_feature    27527. .41619
                     /note="assembly_fragment"
     misc_feature    41720. .98140
                     /note="assembly_fragment"
```

```
     misc_feature    98241. .120308
                     /note="assembly_fragment"
     misc_feature    120409. .160843
                     /note="assembly_fragment"
     misc_feature    160944. .216116
                     /note="assembly_fragment"
     misc_feature    216217. .229102
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             14.0%;  Score 33.4;  DB 2;  Length 229102;
  Best Local Similarity   54.5%;  Pred. No. 44;
  Matches   67;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          50 GCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGA 109
               | | | ||| |  |||  ||||  || | | ||||  |  |  ||  | | | ||||  |
Db      122376 GGCCAGGGGGGCTGTGGATTTCCTTTAGTTTCAGAGCACTCAGACTGCCAAATCCCTGTA 122435

Qy         110 CCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGC 169
               |  |    |  |||| |  |    |||| |   ||| ||  |    |||  ||| |||||
Db      122436 CTCTGAGTTGTTATTTTCTGGGCCCAAAAAAAAAAAAAAAAAAAAAAAAGGGCTGAGTGGC 122495

Qy         170 AAG 172
                ||
Db      122496 CAG 122498


Search completed: September 18, 2004, 12:08:37
Job time : 1148.55 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 09:40:06 ; Search time 257.309 Seconds
(without alignments)
3945.922 Million cell updates/sec

Title: US-10-614-282-1
Perfect score: 239
Sequence: 1 atcagtcacgacttggtaag.....tgtggttccggctggacaat 239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	239	100.0	18920	4	ABL02306	Abl02306 Drosophil
c	2	32.8	13.7	727	6	ABQ70202	Abq70202 Listeria
c	3	32	13.4	3250	9	ADB68937	Adb68937 C. neofor
c	4	31.8	13.3	17468	6	ABA01441	Aba01441 Streptoco
	5	31.6	13.2	969	7	ACA36437	Aca36437 Prokaryot
c	6	31.6	13.2	110000	6	ABA03041_22	Continuation (23 o
	7	31.2	13.1	11638	4	ABL04264	Abl04264 Drosophil
	8	31	13.0	536	6	ABQ30441	Abq30441 Oligonucl
c	9	31	13.0	536	6	ABQ30440	Abq30440 Oligonucl
c	10	31	13.0	6061	4	AAS45335	Aas45335 Chemicall
	11	30.8	12.9	391	2	AAQ60282	Aaq60282 Human bra
	12	30.8	12.9	1625	4	ABL20185	Abl20185 Drosophil
	13	30.8	12.9	1738	4	ABL14353	Abl14353 Drosophil
c	14	30.8	12.9	106286	6	ABS55320_4	Continuation (5 of
c	15	30.8	12.9	110000	6	ABS55320_2	Continuation (3 of
c	16	30.8	12.9	341511	6	ABS55200	Abs55200 Genomic D
	17	30.2	12.6	1360	5	ABA21393	Aba21393 Human ner
	18	30.2	12.6	1360	5	ABA21394	Aba21394 Human ner
	19	30.2	12.6	2000	7	ADA71938	Ada71938 Rice gene
c	20	30.2	12.6	167343	6	ABL64403	Abl64403 Stomach c
c	21	30.2	12.6	167343	6	ABL67239	Abl67239 Thyroid c
c	22	30	12.6	873	6	ABQ68462	Abq68462 Listeria
	23	30	12.6	2377	6	ABQ70841	Abq70841 Listeria
	24	29.8	12.5	438	5	ABA20697	Aba20697 Human ner
	25	29.8	12.5	349980	5	AAH41224	Aah41224 Pyrococcu
	26	29.6	12.4	591	7	ACF73872	Acf73872 Staphyloc
	27	29.6	12.4	1225	2	AAQ73396	Aaq73396 CviJI ORF
c	28	29.6	12.4	1472	2	AAV74866	Aav74866 Staphyloc
	29	29.6	12.4	5496	2	AAQ73395	Aaq73395 CviJI cod
	30	29.6	12.4	110000	6	ABA03041_21	Continuation (22 o
	31	29.4	12.3	6594	3	AAA70287	Aaa70287 Plasmodiu
c	32	29.4	12.3	43064	7	ACC80906	Acc80906 Human CYP
	33	29.2	12.2	468	3	AAC35767	Aac35767 Zea mays
c	34	29.2	12.2	1254	6	ABK35684	Abk35684 cDNA sequ
c	35	29.2	12.2	6200	5	AAS72544	Aas72544 DNA encod
	36	29	12.1	412	7	ABX46645	Abx46645 Bovine ES
c	37	29	12.1	439	5	ABV15324	Abv15324 Human pro
c	38	29	12.1	3863	7	ACA23730	Aca23730 Prokaryot
	39	28.8	12.1	777	7	ACA30362	Aca30362 Prokaryot
	40	28.8	12.1	1074	7	ACA52220	Aca52220 Prokaryot
	41	28.8	12.1	349901	9	ADC86940	Adc86940 Human GPC
	42	28.8	12.1	349938	9	ADC87621	Adc87621 Human GPC
c	43	28.6	12.0	409	4	AAS37706	Aas37706 Novel hum
	44	28.6	12.0	768	4	AAH52955	Aah52955 S. epider
c	45	28.6	12.0	923	7	ABZ51987	Abz51987 Aspergill

ALIGNMENTS

RESULT 1
ABL02306
ID ABL02306 standard; cDNA; 18920 BP.
XX
AC ABL02306;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1400.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB58203.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 1400; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly

```
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 18920 BP; 5734 A; 4166 C; 3925 G; 5095 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 239;  DB 4;  Length 18920;
  Best Local Similarity  100.0%;  Pred. No. 2e-71;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 821

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 881

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 941

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 1000


RESULT 2
ABQ70202/c
ID   ABQ70202 standard; DNA; 727 BP.
XX
AC   ABQ70202;
XX
DT   29-AUG-2003  (revised)
DT   29-AUG-2002  (first entry)
XX
DE   Listeria monocytogenes 4b contig DNA sequence #144.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria monocytogenes ATCC 19115.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR003061.
XX
PR   04-OCT-2000; 2000FR-00012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators.
XX
PS   Claim 14; SEQ ID NO 3015; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences (ABQ67188-
CC   ABQ71212) from Listeria sp. The sequences are useful as probes and
CC   primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of gene
CC   expression. Proteins encoded by the nucleic acid sequences can be used to
CC   screen for compounds that modulate gene expression, replication and
CC   pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in anti-
CC   Listeria vaccines. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
CC   on 29-AUG-2003 to standardise OS field)
XX
SQ   Sequence 727 BP; 225 A; 106 C; 164 G; 231 T; 0 U; 1 Other;

  Query Match            13.7%;  Score 32.8;  DB 6;  Length 727;
  Best Local Similarity  64.5%;  Pred. No. 0.92;
  Matches   49;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy         88 AACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGA 147
              |||| |  | ||||| ||||| ||||  ||||  |||| |  | | |||| ||  ||  |
Db        214 AACCAAAATTGATAGGCCTCGTCCGTATCGTATAATTCCTTCACACCAAAAAGTAAACTA 155

Qy        148 ACTATTTCAAGAACTG 163
                 |   ||||| |||
Db        154 GACAAAACAAGATCTG 139


RESULT 3
ADB68937/c
ID   ADB68937 standard; DNA; 3250 BP.
XX
AC   ADB68937;
XX
DT   04-DEC-2003  (first entry)
XX
DE   C. neoformans genomic DNA sequence SEQ ID NO:64.
XX
KW   ds; gene; fungicide; gene therapy; infection.
XX
OS   Cryptococcus neoformans.
XX
PN   WO2003052076-A2.
XX
PD   26-JUN-2003.
XX
PF   17-DEC-2002; 2002WO-US040225.
XX
PR   17-DEC-2001; 2001US-0341261P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zamudio C,  Eroshkin AM;
XX
DR   WPI; 2003-533017/50.
DR   P-PSDB; ADB70020.
XX
PT   New nucleic acid, useful for preparing a composition for treating an
PT   infection caused by Cryptococcus neoformans.
XX
PS   Claim 3; SEQ ID NO 64; 136pp; English.
XX
CC   The invention relates to a novel purified or isolated Cryptococcus
CC   neoformans nucleic acid molecule comprising a sequence encoding a
CC   polypeptide comprising a sequence not given in the specification. A
CC   polynucleotide of the invention has fungicide activity, and may have a
CC   use in gene therapy. The nucleic acid is useful for preparing a
CC   composition for treating an infection caused by Cryptococcus neoformans.
CC   The present sequence represents a C. neoformans sequence of the
CC   invention. Note: The sequence data for this patent is not represented in
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ   Sequence 3250 BP; 742 A; 827 C; 884 G; 797 T; 0 U; 0 Other;

  Query Match            13.4%;  Score 32;  DB 9;  Length 3250;
  Best Local Similarity  52.2%;  Pred. No. 3.3;
  Matches   71;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          7 CACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGT 66
              || | | | |  | | | |   |||||  ||  ||||  | | |||| ||||     ||  |
Db        395 CAGGGCATCGGCATCCCCCTCACAGCAGATCTACGTCCTAACCGCCGACGGTTCACGCCT 336

Qy         67 TTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCT 126
```

```
Db        335 CTTCCTCCTCGATCCTACAAGACCGAACGAGGAAGAACCTCCACCATACGCATCTTTCCC 276

Qy        127 TAGAAAGCAAACAGCT 142
              || | | || |  ||
Db        275 TATACAACATATCCCT 260


RESULT 4
ABA01441/c
ID   ABA01441 standard; DNA; 17468 BP.
XX
AC   ABA01441;
XX
DT   21-FEB-2002  (first entry)
XX
DE   Streptococcus thermophilus eps3 operon #2.
XX
KW   Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
KW   ds.
XX
OS   Streptococcus thermophilus.
XX
PN   WO200179500-A2.
XX
PD   25-OCT-2001.
XX
PF   18-APR-2001; 2001WO-FR001199.
XX
PR   18-APR-2000; 2000FR-00004972.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
PA   (DANO-) CIE DANONE SA GERVAIS.
PA   (RHOD ) RHODIA CHIM.
XX
PI   Rallu F,  Besancon-Yoshpe I,  Fremaux C,  Mengaud J,  Renault P;
XX
DR   WPI; 2002-017616/02.
XX
PT   New nucleic acid fragments containing exopolysaccharide operon, useful
PT   e.g. for increasing exopolysacchride synthesis in lactic acid bacteria.
XX
PS   Claim 9; Page 75-80; 144pp; French.
XX
CC   The present sequence is an eps operon from Streptococcus thermophilus.
CC   Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC   synthesis. The operon is useful for producing chimeric eps operons, for
CC   optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC   mouth feel and rheological properties to fermented food products (e.g.
CC   yoghurt). They function as thickeners, to provide free-flowing and creamy
CC   texture, and may also have biological activities beneficial to health
XX
SQ   Sequence 17468 BP; 5871 A; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;

  Query Match            13.3%;  Score 31.8;  DB 6;  Length 17468;
  Best Local Similarity  56.1%;  Pred. No. 7.9;
  Matches   60;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         85 TAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAA 144
              ||||| ||||| | |||| | || ||  |  | | |||| |    ||| | | ||| ||
Db       7627 TAAAAACCACGACAATAGTCTTCAACAATATCTTGATATCCGACCAAATCGACCAGTTAT 7568

Qy        145 AGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGATA 191
              ||  |||| |      || | | |    |    |||| || |||
Db       7567 CGATATATTCAACATCGAGTCTTACCACCTCATCAAAGTTGGTAATA 7521


RESULT 5
ACA36437
ID   ACA36437 standard; DNA; 969 BP.
XX
AC   ACA36437;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #18094.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Listeria monocytogenes.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU32567.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 24307; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 969 BP; 320 A; 166 C; 190 G; 293 T; 0 U; 0 Other;

  Query Match            13.2%;  Score 31.6;  DB 7;  Length 969;
  Best Local Similarity  65.7%;  Pred. No. 2.7;
  Matches   46;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;
```

```
Qy        108 GACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTG 167
              || |   | | || ||| |||||||   |  | |||||| | || || || ||  ||
Db          7 GAACAAAGAGAAAAATTATTAGAAATTCATAATCTAAAGCAATACTTTAACAAAGGTACT 66

Qy        168 GCAAGTGAAG 177
              ||||||||||
Db         67 GCAAGTGAAG 76
```

RESULT 6
ABA03041_22/c
Continuation (23 of 30) of ABA03041 from base 2200001 (Listeria monocytogenes EGD-e genc
WP Sequence split into 30 fragments LOCUS ABA03041 Accession Aba03041

WP	Fragment Name	Begin	End
WP	ABA03041_00	1	110000
WP	ABA03041_01	100001	210000
WP	ABA03041_02	200001	310000
WP	ABA03041_03	300001	410000
WP	ABA03041_04	400001	510000
WP	ABA03041_05	500001	610000
WP	ABA03041_06	600001	710000
WP	ABA03041_07	700001	810000
WP	ABA03041_08	800001	910000
WP	ABA03041_09	900001	1010000
WP	ABA03041_10	1000001	1110000
WP	ABA03041_11	1100001	1210000
WP	ABA03041_12	1200001	1310000
WP	ABA03041_13	1300001	1410000
WP	ABA03041_14	1400001	1510000
WP	ABA03041_15	1500001	1610000
WP	ABA03041_16	1600001	1710000
WP	ABA03041_17	1700001	1810000
WP	ABA03041_18	1800001	1910000
WP	ABA03041_19	1900001	2010000
WP	ABA03041_20	2000001	2110000
WP	ABA03041_21	2100001	2210000
WP	ABA03041_22	2200001	2310000
WP	ABA03041_23	2300001	2410000
WP	ABA03041_24	2400001	2510000
WP	ABA03041_25	2500001	2610000
WP	ABA03041_26	2600001	2710000
WP	ABA03041_27	2700001	2810000
WP	ABA03041_28	2800001	2910000
WP	ABA03041_29	2900001	2944528

Query Match 13.2%; Score 31.6; DB 6; Length 110000;
Best Local Similarity 65.7%; Pred. No. 20;
Matches 46; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy        108 GACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTG 167
              || |   | | || ||| |||||||   |  | |||||| | || || || ||  ||
Db      81200 GAACAAAGAGAAAAATTATTAGAAATTCATAATCTAAAGCAATACTTTAACAAAGGTACT 81141

Qy        168 GCAAGTGAAG 177
              ||||||||||
Db      81140 GCAAGTGAAG 81131
```

RESULT 7
ABL04264
ID ABL04264 standard; cDNA; 11638 BP.
XX
AC ABL04264;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 7274.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB60161.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 7274; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11638 BP; 3301 A; 2602 C; 2428 G; 3307 T; 0 U; 0 Other;

Query Match 13.1%; Score 31.2; DB 4; Length 11638;
Best Local Similarity 51.4%; Pred. No. 11;
Matches 72; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy         85 TAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAA 144
              | | || ||| ||| | || |    | |   | || |   | | |||||  | | ||
Db       9470 TCACACACACATCGCTTTCCACCTCAAGCCCTATTCTAATGAAATATAGCAATTACCAAA 9529

Qy        145 AGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGATACACCGGTTATATC 204
              | | ||  ||  |||| || | | |||| |  | | | | | | ||   |  | ||
Db       9530 GATAATTTTCAAATTACTGCATGCCCATTGAATGTGATTCAATTAGAGACTCCTGTTTTC 9589

Qy        205 GGAGTGGCGAGAAAGTGTGG 224
                |    || |||  ||||
Db       9590 CCAAGCAAGACAAAAGGTGG 9609
```

RESULT 8
ABQ30441
ID ABQ30441 standard; DNA; 536 BP.
XX
AC ABQ30441;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17032.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX

PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 536 BP; 189 A; 188 C; 65 G; 94 T; 0 U; 0 Other;

Query Match 13.0%; Score 31; DB 6; Length 536;
Best Local Similarity 64.8%; Pred. No. 3.4;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy         34 CGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA 93
              |||||||| || ||||  | ||| || |||  | ||||| ||| |   | | ||| ||
Db        193 CGTCGTCGCCGCCATCCTCGACGCGACTCGCTTCTTTCGATTCTACCTAAAAAAATCCAC 252

Qy         94 CGTCGATAGCC 104
              || |   |||
Db        253 CGCCATCCGCC 263
```

RESULT 9
ABQ30440/c
ID ABQ30440 standard; DNA; 536 BP.
XX
AC ABQ30440;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17031.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 536 BP; 94 A; 65 C; 188 G; 189 T; 0 U; 0 Other;

Query Match 13.0%; Score 31; DB 6; Length 536;
Best Local Similarity 64.8%; Pred. No. 3.4;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy         34 CGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA 93
              |||||||| || ||||  | ||| || |||  | ||||| ||| |   | | ||| ||
Db        344 CGTCGTCGCCGCCATCCTCGACGCGACTCGCTTCTTTCGATTCTACCTAAAAAAATCCAC 285

Qy         94 CGTCGATAGCC 104
              || |   |||
Db        284 CGCCATCCGCC 274
```

RESULT 10
AAS45335/c
ID AAS45335 standard; DNA; 6061 BP.
XX
AC AAS45335;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated complementary DNA associated with cell cycle #20.
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
OS Homo sapiens.
XX

PN WO200168911-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002945.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
PS Claim 1; SEQ ID NO 40; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
SQ Sequence 6061 BP; 1558 A; 264 C; 1557 G; 2682 T; 0 U; 0 Other;

Query Match 13.0%; Score 31; DB 4; Length 6061;
Best Local Similarity 64.8%; Pred. No. 9.6;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy         34 CGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA 93
              |||||||| || ||||  | ||| || |||  | ||||| ||| |   | | ||| ||
Db       2164 CGTCGTCGCCGCCATCCTCGACGCGACTCGCTTCTTTCGATTCTACCTAAAAAAATCCAC 2105

Qy         94 CGTCGATAGCC 104
              || |   |||
Db       2104 CGCCATCCGCC 2094
```

RESULT 11
AAQ60282
ID AAQ60282 standard; DNA; 391 BP.
XX
AC AAQ60282;
XX
DT 25-MAR-2003 (revised)
DT 16-MAR-1994 (first entry)
XX
DE Human brain Expressed Sequence Tag EST02279.
XX
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS Homo sapiens.
XX
PN WO9316178-A2.
XX
PD 19-AUG-1993.
XX
PF 12-FEB-1993; 93WO-US001294.
XX
PR 12-FEB-1992; 92US-00837195.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Venter CJ, Adams MD, Moreno RF;
XX
DR WPI; 1993-272882/34.
XX
PT Enriched oligonucleotides and corresp. sequences - used as markers for
PT human genes transcribed in-vivo, facilitate tagging of most human genes.
XX
PS Example 4; Page 310; 500pp; English.
XX
CC The Expressed Sequence Tag was isolated from a human brain cDNA library
CC as part of a large set of ESTs which can be used as markers for human
CC genes transcribed in vivo. They can be used to facilitate tagging of most
CC human genes, for mapping locations of expressed genes on chromosomes, for
CC individual or forensic identification, for mapping locations of disease-
CC associated genes, for identification of tissue type, and for prepn. of
CC antisense sequences, probes and constructs. EST02279 has a "poor" coding
CC probability as evaluated using the coding-region prediction program CRM.
CC See also AAQ59041-Q61440. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 391 BP; 96 A; 94 C; 89 G; 109 T; 0 U; 3 Other;

Query Match 12.9%; Score 30.8; DB 2; Length 391;
Best Local Similarity 54.6%; Pred. No. 3.5;
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy        117 GTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAA 176
              | |  || ||||    ||| |   || |        |||| ||| |   || ||||| ||
Db         33 GCAGCATCCTTACCTTGCAGAGTACTGAGTCCTGGCTTCATGAATTTNATGTCAAGTAAA 92

Qy        177 GGGTAGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGG 224
                ||  ||||| || |   | | || |    ||  ||||||  | ||
Db         93 TGGGTTTTAGTCATCCCTAGTTCATGTGCATGTNCCGAGAAAAAGGGG 140
```

RESULT 12
ABL20185
ID ABL20185 standard; DNA; 1625 BP.
XX
AC ABL20185;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12028.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more

```
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 12028; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1625 BP; 410 A; 413 C; 387 G; 415 T; 0 U; 0 Other;

  Query Match             12.9%;  Score 30.8;  DB 4;  Length 1625;
  Best Local Similarity   58.9%;  Pred. No. 6.4;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy         14 TGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGG 73
              ||||| | |  ||| |||  |||||||  || | | |  | |   ||    ||| | | |
Db        611 TGGTACGTGGTCAGCCAGTTCGTCGTCTCCGCCGTTGGAATCTTCAGCAAGGTTGTGCTG 670

Qy         74 TTCGATACAGATAAAACCCACGTCGATAGC 103
                |   |    ||||| ||| |||| ||| |
Db        671 ATGTTTTTGAATAAACCCCGCGTCTATAAC 700


RESULT 13
ABL14353
ID   ABL14353 standard; cDNA; 1738 BP.
XX
AC   ABL14353;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 37541.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB70250.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 37541; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1738 BP; 454 A; 440 C; 425 G; 419 T; 0 U; 0 Other;

  Query Match             12.9%;  Score 30.8;  DB 4;  Length 1738;
  Best Local Similarity   58.9%;  Pred. No. 6.6;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy         14 TGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGG 73
              ||||| | |  ||| |||  |||||||  || | | |  | |   ||    ||| | | |
Db        724 TGGTACGTGGTCAGCCAGTTCGTCGTCTCCGCCGTTGGAATCTTCAGCAAGGTTGTGCTG 783

Qy         74 TTCGATACAGATAAAACCCACGTCGATAGC 103
                |   |    ||||| ||| |||| ||| |
Db        784 ATGTTTTTGAATAAACCCCGCGTCTATAAC 813


RESULT 14
ABS55320_4/c
Continuation (5 of 5) of ABS55320 from base 400001 (Human transporter protein genomic DN
WP Sequence split into 5 fragments  LOCUS ABS55320 Accession Abs55320
WP     Fragment Name      Begin     End
WP     ABS55320_0             1     110000
WP     ABS55320_1        100001     210000
WP     ABS55320_2        200001     310000
WP     ABS55320_3        300001     410000
WP     ABS55320_4        400001     506286

  Query Match             12.9%;  Score 30.8;  DB 6;  Length 106286;
  Best Local Similarity   50.7%;  Pred. No. 38;
  Matches   74;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         45 TCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCC 104
              |||  ||    |  ||| | |  || ||  |   |   | | ||||| | ||  | ||
Db      38129 TCAATGCAGGCAAGAGCCATCCTTCTCTATGTTAAGAAAAGAAAACCAAAGTTAAAAGGA 38070

Qy        105 CTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGT 164
                |  || | |   |  | || ||  | |||| ||  | ||  |||| ||    | |
Db      38069 TAAAAGAATCTCCTCTTTCACCTAAAATACTAACAACTGCATAAGCATTTTAATTTTTAT 38010

Qy        165 GTGGCAAGTGAAGGGTAGTTAGTGAT 190
                || || ||||      ||| | ||
Db      38009 TAGGGAATTGAAAAAATATTAATTAT 37984


RESULT 15
ABS55320_2/c
Continuation (3 of 5) of ABS55320 from base 200001 (Human transporter protein genomic DN
WP Sequence split into 5 fragments  LOCUS ABS55320 Accession Abs55320
WP     Fragment Name      Begin     End
WP     ABS55320_0             1     110000
WP     ABS55320_1        100001     210000
WP     ABS55320_2        200001     310000
WP     ABS55320_3        300001     410000
WP     ABS55320_4        400001     506286

  Query Match             12.9%;  Score 30.8;  DB 6;  Length 110000;
  Best Local Similarity   50.7%;  Pred. No. 39;
  Matches   74;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         45 TCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCC 104
              |||  ||    |  ||| | |  || ||  |   |   | | ||||| | ||  | ||
Db      10425 TCAATGCAGGCAAGAGCCATCCTTCTCTATGTTAAGAAAAGAAAACCAAAGTTAAAAGGA 10366

Qy        105 CTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGT 164
                |  || | |   |  | || ||  | |||| ||  | ||  |||| ||    | |
Db      10365 TAAAAGAATCTCCTCTTTCACCTAAAATACTAACAACTGCATAAGCATTTTAATTTTTAT 10306

Qy        165 GTGGCAAGTGAAGGGTAGTTAGTGAT 190
```

```
                   || || ||||      ||| | ||
Db      10305 TAGGGAATTGAAAAAATATTAATTAT 10280
```

Search completed: September 18, 2004, 11:29:54
Job time : 261.309 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:14:16 ; Search time 59.8737 Seconds
(without alignments)
2215.217 Million cell updates/sec

Title: US-10-614-282-1
Perfect score: 239
Sequence: 1 atcagtcacgacttggtaag.....tgtggttccggctggacaat 239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	29.6	12.4	1225	1	US-08-181-629A-3	Sequence 3, Appli
c	2	29.6	12.4	1472	4	US-08-956-171E-555	Sequence 555, App
	3	29.6	12.4	5496	1	US-08-181-629A-2	Sequence 2, Appli
	4	28.6	12.0	1731	4	US-09-134-001C-1118	Sequence 1118, Ap
	5	27.8	11.6	423	4	US-09-397-787-196	Sequence 196, App
	6	27.6	11.5	716	3	US-08-998-416-746	Sequence 746, App
c	7	27.4	11.5	1296	4	US-09-543-681A-2060	Sequence 2060, Ap
c	8	27.4	11.5	1650	4	US-09-107-532A-3186	Sequence 3186, Ap
	9	27	11.3	1200	4	US-09-358-082A-27	Sequence 27, Appl
	10	27	11.3	1521	4	US-09-328-352-1397	Sequence 1397, Ap
	11	27	11.3	2026	2	US-08-993-228-3	Sequence 3, Appli
	12	27	11.3	15071	4	US-09-358-082A-29	Sequence 29, Appl
	13	26.8	11.2	505	4	US-09-621-976-15639	Sequence 15639, A
c	14	26.8	11.2	612	4	US-09-716-865-7	Sequence 7, Appli
	15	26.8	11.2	1975	4	US-09-910-174B-3	Sequence 3, Appli
	16	26.8	11.2	1975	4	US-09-620-461-3	Sequence 3, Appli
	17	26.8	11.2	2229	4	US-09-910-174B-1	Sequence 1, Appli
	18	26.8	11.2	2229	4	US-09-620-461-1	Sequence 1, Appli
	19	26.8	11.2	10754	2	US-08-966-958-1	Sequence 1, Appli
	20	26.8	11.2	10754	2	US-09-215-817-1	Sequence 1, Appli
	21	26.8	11.2	10754	3	US-09-342-353-1	Sequence 1, Appli
c	22	26.8	11.2	55216	4	US-09-716-865-23	Sequence 23, Appl
c	23	26.6	11.1	657	4	US-09-489-039A-2668	Sequence 2668, Ap
c	24	26.6	11.1	1128	4	US-09-489-039A-4389	Sequence 4389, Ap
c	25	26.6	11.1	2147	1	US-08-448-600-1	Sequence 1, Appli
c	26	26.6	11.1	3088	3	US-08-937-610-3	Sequence 3, Appli
	27	26.6	11.1	4047	4	US-09-534-407-4	Sequence 4, Appli
	28	26.6	11.1	4047	4	US-09-999-201B-3	Sequence 3, Appli
	29	26.6	11.1	9610	4	US-09-566-921-45	Sequence 45, Appl
c	30	26.6	11.1	640681	4	US-09-790-988-1	Sequence 1, Appli
	31	26.4	11.0	2515	3	US-08-714-918-37	Sequence 37, Appl
	32	26.4	11.0	2515	3	US-09-265-315-37	Sequence 37, Appl
	33	26.4	11.0	2515	3	US-09-265-315-37	Sequence 37, Appl
	34	26.4	11.0	2515	3	US-09-266-417-37	Sequence 37, Appl
	35	26.4	11.0	2515	4	US-09-528-709-37	Sequence 37, Appl
	36	26.4	11.0	2515	4	US-09-527-745-37	Sequence 37, Appl
c	37	26.4	11.0	7762	4	US-08-956-171E-124	Sequence 124, App
	38	26.4	11.0	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	39	26.2	11.0	87350	3	US-08-781-891-79	Sequence 79, Appl
	40	26.2	11.0	87350	4	US-09-618-166-79	Sequence 79, Appl
	41	26.2	11.0	87543	4	US-09-791-211-3	Sequence 3, Appli
c	42	26	10.9	956	4	US-09-641-638-42	Sequence 42, Appl
c	43	26	10.9	956	4	US-09-641-638-43	Sequence 43, Appl
c	44	26	10.9	1001	4	US-09-641-638-620	Sequence 620, App
c	45	26	10.9	1001	4	US-09-641-638-621	Sequence 621, App

ALIGNMENTS

```
RESULT 1
US-08-181-629A-3
; Sequence 3, Application US/08181629A
; Patent No. 5472872
;  GENERAL INFORMATION:
;    APPLICANT:  Swaminathan, Neela
;    APPLICANT:  Van Etten, James
;    APPLICANT:  Mead, David
;    APPLICANT:  Skowron, Piotr
;    TITLE OF INVENTION:  Recombinant CviJI Restriction Endonuclease
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/181,629A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31504
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..33, 55..1128)
US-08-181-629A-3

  Query Match             12.4%;  Score 29.6;  DB 1;  Length 1225;
  Best Local Similarity   64.7%;  Pred. No. 1.1;
  Matches   44;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;
```

```
Qy        131 AAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGAT 190
              |||   ||||  |||||| ||    |||  || |    |||| |||  |||||||||
Db        691 AAGTTCACAGGAAAAGAATTAGAAGAAGTTCTATCGTTCAAGAGAAAAGTAGTTAGTATG 750

Qy        191 ACACCGGT 198
              ||||||| 
Db        751 GCACCGGT 758


RESULT 2
US-08-956-171E-555/c
; Sequence 555, Application US/08956171E
; Patent No. 6593114
;  GENERAL INFORMATION:
;       APPLICANT: Charles Kunsch
;                  Gil H. Choi
;                  Patrick S. Dillon
;                  Craig A. Rosen
;                  Steven C. Barash
;                  Michael R. Fannon
;       TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 5256
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/956,171E
;            FILING DATE: 20-Oct-1997
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/009,861
;            FILING DATE: January 5, 1996
;            APPLICATION NUMBER: 08/781,986
;            FILING DATE: January 3, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mark J. Hyman
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB248P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (240) 314-1224
;            TELEFAX: (301) 309-8439
;  INFORMATION FOR SEQ ID NO: 555:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1472 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 555:
US-08-956-171E-555

  Query Match            12.4%;  Score 29.6;  DB 4;  Length 1472;
  Best Local Similarity   50.0%;  Pred. No. 1.2;
  Matches   74;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         60 GTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTA 119
              ||   || | |   || | | | ||| |  |||| | ||   |||  ||     | ||
Db       1016 GTATCGTGTCTTATTTTGGTGGAAATATCATTCACGGCTATCATCCTGTACACGGACATA 957

Qy        120 ATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGG 179
              ||  |    |   || ||   | |||   | ||||||| ||||  |   || |  |  |
Db        956 CTACACAGTTACGCCATACCAATGAAGGTATTTTTCAAGGACTGCCTCAAAATTTCAATG 897

Qy        180 TAGTTAGTGATACACCGGTTATATCGGA 207
                 || |  || ||    |  | ||  | ||
Db        896 TAATGCGTTATCATTCATTAATTGCTGA 869


RESULT 3
US-08-181-629A-2
; Sequence 2, Application US/08181629A
; Patent No. 5472872
;  GENERAL INFORMATION:
;    APPLICANT:  Swaminathan, Neela
;    APPLICANT:  Van Etten, James
;    APPLICANT:  Mead, David
;    APPLICANT:  Skowron, Piotr
;    TITLE OF INVENTION:  Recombinant CviJI Restriction Endonuclease
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/181,629A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31504
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5496 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-181-629A-2

  Query Match            12.4%;  Score 29.6;  DB 1;  Length 5496;
  Best Local Similarity   64.7%;  Pred. No. 2.2;
  Matches   44;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        131 AAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGAT 190
              |||   ||||  |||||| ||    |||  || |    |||| |||  |||||||||
Db       1759 AAGTTCACAGGAAAAGAATTAGAAGAAGTTCTATCGTTCAAGAGAAAAGTAGTTAGTATG 1818

Qy        191 ACACCGGT 198
              ||||||| 
Db       1819 GCACCGGT 1826


RESULT 4
US-09-134-001C-1118
; Sequence 1118, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
```

```
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1118
;   LENGTH: 1731
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1118

  Query Match            12.0%;  Score 28.6;  DB 4;  Length 1731;
  Best Local Similarity   59.0%;  Pred. No. 2.9;
  Matches   49;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         77 GATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAA 136
              || ||| || ||| | |     | ||| | | || ||| |   | ||   ||     |||
Db       1618 GAAACACATCAAAGCGAAAAACAAAGCACACAACAGTCTCAACAAATAGATATTGCTCAA 1677

Qy        137 ACAGCTAAAGAACTATTTCAAGA 159
              | |||||||||  |||||   ||
Db       1678 AAAGCTAAAGACTTATTTGGTGA 1700


RESULT 5
US-09-397-787-196
; Sequence 196, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: King, Gordon E.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
;  TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
;  FILE REFERENCE: 210121.466C2
;  CURRENT APPLICATION NUMBER: US/09/397,787
;  CURRENT FILING DATE: 1999-09-16
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 196
;   LENGTH: 423
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-397-787-196

  Query Match            11.6%;  Score 27.8;  DB 4;  Length 423;
  Best Local Similarity   57.5%;  Pred. No. 2.8;
  Matches   50;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy         99 ATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAG 158
              ||  |||||  ||  | ||  |||||| ||  |||  ||| |   |||  |||    |  ||
Db         66 ATTCCCCTCCCCCATAGCAAAATATTATTTAAAAAAAAAAAAAACAAAAAACAGGGTGGAG 125

Qy        159 AACTGTGTGGCAAGTGAAGGGTAGTTA 185
              |   |  | | |||  | |  || | ||
Db        126 AGGAGGATAGGAAGGGGACAGTTGATA 152


RESULT 6
US-08-998-416-746
; Sequence 746, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  746:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  716 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1484UP
US-08-998-416-746

  Query Match            11.5%;  Score 27.6;  DB 3;  Length 716;
  Best Local Similarity   55.1%;  Pred. No. 4.3;
  Matches   54;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy         16 GTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTT 75
              ||| | ||  | | | ||| |  |||  ||| ||   ||| ||   || |  | | ||
Db        454 GTACGAGCCGATGGTGAACGACTACGTGCTCAACGTGGACGAGAACTGTTTGATACAGTA 513

Qy         76 CGATACAGATAAAACCCACGTCGATAGCCCTCGACCGT 113
              |||   | | ||||  ||  | ||   |  ||| | ||
Db        514 CGAGCTATACAAAAGGCAGCTGGAGCACAATCGGCAGT 551


RESULT 7
US-09-543-681A-2060/c
; Sequence 2060, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABII
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2060
;   LENGTH: 1296
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-2060

  Query Match            11.5%;  Score 27.4;  DB 4;  Length 1296;
  Best Local Similarity   55.9%;  Pred. No. 6.8;
  Matches   52;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;
```

```
Qy         66 TTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTC 125
              ||  |  |  ||||| | ||||| ||  | || | ||||  ||      |   ||||
Db       1235 TTGCTATGAGCGATAGACATAAAGCCTGCTTCAAAAGCCGACGGAGCGAGGTAAATACGT 1176

Qy        126 TTAGAAAGCAAACAGCTAAAGAACTATTTCAAG 158
              ||    | ||   ||   |||||||| ||| |||
Db       1175 TTTTCTAACATTAAGTGGAAGAACTGTTTAAAG 1143


RESULT 8
US-09-107-532A-3186/c
; Sequence 3186, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 3186:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1650 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: circular
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...1650
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3186:
US-09-107-532A-3186

  Query Match             11.5%;  Score 27.4;  DB 4;  Length 1650;
  Best Local Similarity   59.7%;  Pred. No. 7.6;
  Matches   46;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy         33 ACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCC 92
              |||||  |    | | |||| |||  ||||      |   |||| | ||| ||||||||  ||
Db        206 ACGTCATAACCATTATCGACAAATGGAGATTGATAGATCGGATTGATCCAGATAAAGCCG 147

Qy         93 ACGTCGATAGCCCTCGA 109
              |   | | | |  || |
```

```
Db        146 ATACCAAGATCTTTCAA 130


RESULT 9
US-09-358-082A-27
; Sequence 27, Application US/09358082A
; Patent No. 6589784
; GENERAL INFORMATION:
;  APPLICANT:  Antoniou, Michael
;  APPLICANT:  Crombie, Robert
;  TITLE OF INVENTION:  A Polynucleotide
;  FILE REFERENCE:  CACO0056
;  CURRENT APPLICATION NUMBER: US/09/358,082A
;  CURRENT FILING DATE:  1999-07-21
;  PRIOR APPLICATION NUMBER:  GB 9815879.3
;  PRIOR FILING DATE:  1998-07-21
;  PRIOR APPLICATION NUMBER:  US 60/107688
;  PRIOR FILING DATE:  1998-11-09
;  PRIOR APPLICATION NUMBER:  GB 9906712.6
;  PRIOR FILING DATE:  1999-03-23
;  PRIOR APPLICATION NUMBER:  US 60/127410
;  PRIOR FILING DATE:  1999-04-01
;  PRIOR APPLICATION NUMBER:  GB 9909494.8
;  PRIOR FILING DATE:  1999-04-23
;  PRIOR APPLICATION NUMBER:  US 60/134016
;  PRIOR FILING DATE:  1999-05-12
;  NUMBER OF SEQ ID NOS:  29
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 1200
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-358-082A-27

  Query Match             11.3%;  Score 27;  DB 4;  Length 1200;
  Best Local Similarity   54.5%;  Pred. No. 9.1;
  Matches   54;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         71 CGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGA 130
              | | |  |||||| |  |||||||| | |  ||    |  ||  | |  | |  |||| ||
Db        362 CAGCTAAATACAGTTGCAACCCACATTGCAAGTGGTATACATTGGAATGAGGGTCTTTGA 421

Qy        131 AAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGC 169
              |   |||     ||||  |  |||  ||  | ||  | ||
Db        422 AGTTAAATCCTTAAACCATGATTCAAACCATTGCTTAGC 460


RESULT 10
US-09-328-352-1397
; Sequence 1397, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1397
;   LENGTH: 1521
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-1397

  Query Match             11.3%;  Score 27;  DB 4;  Length 1521;
  Best Local Similarity   60.0%;  Pred. No. 10;
  Matches   45;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy        115 GCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTG 174
              | | ||||| |||| | ||| ||  |||| ||||| |||      |||  | |||| |
Db       1296 GTGGAATATGCTTAAAGAGCGAAATGCTATAGAACAATTGGGTCAACATTTTGGCCATGA 1355
```

```
Qy        175 AAGGGTAGTTAGTGA 189
              | | |   ||||
Db       1356 TTTGTTTGAGTGTGA 1370


RESULT 11
US-08-993-228-3
; Sequence 3, Application US/08993228
; Patent No. 5976838
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  LaVallie, Edward R.
;    APPLICANT:  Racie, Lisa A.
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    APPLICANT:  Agostino, Michael J.
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/993,228
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sprunger, Suzanne A.
;      REGISTRATION NUMBER:  41,323
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8284
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2026 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-993-228-3

  Query Match             11.3%;  Score 27;  DB 2;  Length 2026;
  Best Local Similarity   53.3%;  Pred. No. 12;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy         63 GTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATA 122
              |||  ||||  ||| ||| |||   | ||||  |  | | |    |    ||  |  | ||
Db       1181 GTGCATTTCACTTCCATAAAGACCCAGCCCAAGCTGCTGGGAACCATGTGTTCCTGAGTA 1240

Qy        123 TTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGC 169
              ||||  |||    |||||  | | |   |  |   |||  |  |||  |
Db       1241 TTCTCAGAGGTTAAACAAGTGACAAGTGAGCTTCTGAAATTAGTGTC 1287


RESULT 12
US-09-358-082A-29
; Sequence 29, Application US/09358082A
; Patent No. 6589784
; GENERAL INFORMATION:
;  APPLICANT:  Antoniou, Michael
;  APPLICANT:  Crombie, Robert
;  TITLE OF INVENTION:  A Polynucleotide
;  FILE REFERENCE:  CACO0056
;  CURRENT APPLICATION NUMBER: US/09/358,082A
;  CURRENT FILING DATE:  1999-07-21
;  PRIOR APPLICATION NUMBER:  GB 9815879.3
;  PRIOR FILING DATE:  1998-07-21
;  PRIOR APPLICATION NUMBER:  US 60/107688
;  PRIOR FILING DATE:  1998-11-09
;  PRIOR APPLICATION NUMBER:  GB 9906712.6
;  PRIOR FILING DATE:  1999-03-23
;  PRIOR APPLICATION NUMBER:  US 60/127410
;  PRIOR FILING DATE:  1999-04-01
;  PRIOR APPLICATION NUMBER:  GB 9909494.8
;  PRIOR FILING DATE:  1999-04-23
;  PRIOR APPLICATION NUMBER:  US 60/134016
;  PRIOR FILING DATE:  1999-05-12
;  NUMBER OF SEQ ID NOS:  29
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 29
;   LENGTH: 15071
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-358-082A-29

  Query Match             11.3%;  Score 27;  DB 4;  Length 15071;
  Best Local Similarity   54.5%;  Pred. No. 32;
  Matches   54;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         71 CGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGA 130
              | | |  |||||| |  ||||||| | |  ||  |  ||  | |  | |   |||| ||
Db      13843 CAGCTAAATACAGTTGCAACCCACATTGCAAGTGGTATACATTGGAATGAGGGTCTTTGA 13902

Qy        131 AAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGC 169
              |   |||    ||||  |  |||  ||  | ||  | ||
Db      13903 AGTTAAATCCTTAAACCATGATTCAAACCATTGCTTAGC 13941


RESULT 13
US-09-621-976-15639
; Sequence 15639, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 15639
;   LENGTH: 505
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-15639

  Query Match             11.2%;  Score 26.8;  DB 4;  Length 505;
  Best Local Similarity   14.6%;  Pred. No. 7;
  Matches   19;  Conservative   62;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         10 GACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTT 69
              | : |:| : | : | : : :|::::: |:  ::       :::  :| : :| :::|:|
Db         39 GKYGTSGGRCGGSRGWMCSKWGYRYSYWGYKWKSWSKKMKKYSGKMGTSKSTRKYRTYTS 98

Qy         70 TCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAG 129
              :|::|||:    ::::::::     : :::::  :|  :: ::: |:  | :|: ::   :
Db         99 KCRKTTCKYRGWWSWKWRMWKRRKMYYYRMKCYSCASYSYSYRRCRKYTGMTRGWYGCKK 158

Qy        130 AAAGCAAACA 139
              :: ::  :::
Db        159 RMCKSSTRYM 168
```

```
RESULT 14
US-09-716-865-7/c
; Sequence 7, Application US/09716865
; Patent No. 6548292
; GENERAL INFORMATION:
;  APPLICANT: Bramucci, Micheal G
;  APPLICANT:  Nagarajan, Vasantha
;  APPLICANT:  Chen, Mario W.
;  TITLE OF INVENTION: Bacterial Plasmid Having Genes Encoding Enzymes for the
;  TITLE OF INVENTION:  Degradation of Aromatic Compounds
;  FILE REFERENCE: BC1016 US NA
;  CURRENT APPLICATION NUMBER: US/09/716,865
;  CURRENT FILING DATE:  2000-11-26
;  PRIOR APPLICATION NUMBER: 60/167,062
;  PRIOR FILING DATE: 1999-11-23
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Pseudomonas CT14
US-09-716-865-7

  Query Match             11.2%;  Score 26.8;  DB 4;  Length 612;
  Best Local Similarity   53.9%;  Pred. No. 7.7;
  Matches   55;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         38 GTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTC 97
              ||  |||||| |  || |||   || || ||||  | ||||    ||  ||  | | |||
Db        129 GTAATCGTCACCAGCATCGGTTTTCTTGGTTTTGAGATCGACCTTGACCAAGTCAAAGTC 70

Qy         98 GATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACA 139
              | | |       ||  | | ||| |     || ||| | ||
Db         69 CAGACCGGCTTCGCGCAGGGCAATGTGGGGTGACAGCGAGCA 28


RESULT 15
US-09-910-174B-3
; Sequence 3, Application US/09910174B
; Patent No. 6630575
; GENERAL INFORMATION:
;  APPLICANT: Coyle, Anthony J.
;  APPLICANT:  Fraser, Christopher C.
;  APPLICANT:  Manning, Stephen
;  TITLE OF INVENTION: B7-H2 Molecules, No. 6630575el Members of the B7
;  TITLE OF INVENTION:  Family and Uses Thereof
;  FILE REFERENCE: 35800/236924
;  CURRENT APPLICATION NUMBER: US/09/910,174B
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 09/620,461
;  PRIOR FILING DATE: 2000-07-20
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1975
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (70)...(618)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1975)
;   OTHER INFORMATION: B7-H2 Short
US-09-910-174B-3

  Query Match             11.2%;  Score 26.8;  DB 4;  Length 1975;
  Best Local Similarity   59.0%;  Pred. No. 14;
  Matches   46;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              |||||||  ||||  |  |   ||||| || |  | | |||   | || |||||||  ||
Db        553 TATTCTTCAAAAGACACAACAAAAAGACCTGTCACCACAACAAAGAGGGAAGTGAACAGT 612

Qy        181 AGTTAGTGATACACCGGT 198
                  |   |||  |   |||
Db        613 GCTATCTGAACCTGTGGT 630


Search completed: September 18, 2004, 13:18:58
Job time : 61.8737 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 12:08:47 ; Search time 263.741 Seconds
(without alignments)
4573.952 Million cell updates/sec

Title: US-10-614-282-1
Perfect score: 239
Sequence: 1 atcagtcacgacttggtaag.....tgtggttccggctggacaat 239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	239	100.0	239	17	US-10-614-282-1	Sequence 1, Appli
c 2	32.8	13.7	727	16	US-10-398-221-3015	Sequence 3015, Ap
3	32.8	13.7	917	13	US-10-424-599-14250	Sequence 14250, A
c 4	32.8	13.7	2731748	17	US-10-297-465A-1	Sequence 1, Appli
c 5	32	13.4	3250	16	US-10-320-797-64	Sequence 64, Appl
6	31.6	13.2	969	13	US-10-282-122A-24307	Sequence 24307, A
c 7	31.4	13.1	661	17	US-10-767-701-1632	Sequence 1632, Ap
8	31.4	13.1	1205	13	US-10-390-184-2	Sequence 2, Appli
c 9	31.4	13.1	1207	17	US-10-437-963-35457	Sequence 35457, A
10	31.4	13.1	1238	13	US-10-390-184-3	Sequence 3, Appli
11	31.4	13.1	158405	15	US-10-175-523-86	Sequence 86, Appl
c 12	31	13.0	6061	15	US-10-239-676-40	Sequence 40, Appl
13	30.2	12.6	498	17	US-10-437-963-92822	Sequence 92822, A
14	30.2	12.6	616	13	US-10-027-632-203363	Sequence 203363,
15	30.2	12.6	616	16	US-10-027-632-203363	Sequence 203363,
c 16	30.2	12.6	693	13	US-10-027-632-32043	Sequence 32043, A
c 17	30.2	12.6	693	16	US-10-027-632-32043	Sequence 32043, A
c 18	30.2	12.6	167343	9	US-09-962-436-281	Sequence 281, App
c 19	30.2	12.6	167343	9	US-09-964-824A-273	Sequence 273, App
20	30	12.6	676	17	US-10-021-323-9249	Sequence 9249, Ap
c 21	30	12.6	873	16	US-10-398-221-1275	Sequence 1275, Ap
22	30	12.6	2377	16	US-10-398-221-3654	Sequence 3654, Ap
c 23	30	12.6	2613	15	US-10-171-319-15	Sequence 15, Appl
c 24	29.8	12.5	1563	16	US-10-369-493-45304	Sequence 45304, A
c 25	29.6	12.4	1472	8	US-08-781-986A-555	Sequence 555, App
c 26	29.6	12.4	1472	13	US-10-329-624-555	Sequence 555, App
27	29.6	12.4	2437	13	US-10-424-599-28760	Sequence 28760, A
28	29.2	12.2	475	13	US-10-425-114-17139	Sequence 17139, A
c 29	29.2	12.2	1254	10	US-09-822-846-75	Sequence 75, Appl
30	29.2	12.2	2090	13	US-10-425-114-2825	Sequence 2825, Ap
31	29.2	12.2	2188	17	US-10-767-701-15426	Sequence 15426, A
32	29.2	12.2	2236	13	US-10-425-114-5797	Sequence 5797, Ap
33	29	12.1	412	9	US-09-960-352-11810	Sequence 11810, A
c 34	29	12.1	3863	13	US-10-282-122A-11600	Sequence 11600, A
c 35	28.8	12.1	442	13	US-10-424-599-59524	Sequence 59524, A
36	28.8	12.1	777	13	US-10-282-122A-18232	Sequence 18232, A
37	28.8	12.1	1074	13	US-10-282-122A-40090	Sequence 40090, A
38	28.8	12.1	822900	16	US-10-292-798-1393	Sequence 1393, Ap
c 39	28.6	12.0	409	10	US-09-803-719-764	Sequence 764, App
40	28.6	12.0	1608	9	US-09-816-028A-24	Sequence 24, Appl
41	28.6	12.0	1608	15	US-10-303-161-24	Sequence 24, Appl
42	28.6	12.0	1608	15	US-10-303-118-24	Sequence 24, Appl
43	28.6	12.0	1608	15	US-10-303-128-24	Sequence 24, Appl
44	28.6	12.0	1608	15	US-10-303-134-24	Sequence 24, Appl
45	28.6	12.0	1608	15	US-10-303-162-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-10-614-282-1
; Sequence 1, Application US/10614282
; Publication No. US20040082034A1
; GENERAL INFORMATION:
; APPLICANT: LEE, DUNG-FANG
; APPLICANT: JUANG, JYH-LYH
; TITLE OF INVENTION: INTERNAL RIBOSOME ENTRY SITE OF THE LABIAL GENE FOR
; TITLE OF INVENTION: PROTEIN EXPRESSION
; FILE REFERENCE: 08842.0001-00000
; CURRENT APPLICATION NUMBER: US/10/614,282
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: 60/394,270
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Drosophila labial
US-10-614-282-1

Query Match 100.0%; Score 239; DB 17; Length 239;
Best Local Similarity 100.0%; Pred. No. 4e-73;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
```

```
Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
```

RESULT 2
US-10-398-221-3015/c
; Sequence 3015, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3015
; LENGTH: 727
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3015

Query Match 13.7%; Score 32.8; DB 16; Length 727;
Best Local Similarity 64.5%; Pred. No. 1;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy         88 AACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGA 147
              |||| |  | ||||| ||||| |||| |||| |||| |  | | |||| ||  ||  |
Db        214 AACCAAAATTGATAGGCCTCGTCCGTATCGTATAATTCCTTCACACCAAAAAGTAAACTA 155

Qy        148 ACTATTTCAAGAACTG 163
                 |  |||||  |||
Db        154 GACAAAACAAGATCTG 139
```

RESULT 3
US-10-424-599-14250
; Sequence 14250, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 14250
; LENGTH: 917
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112876C.1
US-10-424-599-14250

Query Match 13.7%; Score 32.8; DB 13; Length 917;
Best Local Similarity 61.9%; Pred. No. 1.2;
Matches 52; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy        124 TCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGT 183
              |||||| || |  ||  ||| ||||    ||| |||   || ||  ||||   | | |||
Db        246 TCTTAGCAAACGTACTACTACAGAAAGGCTTCGAGAGGAGTTTGCAAAGTTTGGTGAAGT 305

Qy        184 TAGTGATACACCGGTTATATCGGA 207
              |  | || ||  |||| || | ||
Db        306 TGTTCATGCAAGGGTTGTAACTGA 329
```

RESULT 4
US-10-297-465A-1/c
; Sequence 1, Application US/10297465A
; Publication No. US20040142413A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medianis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465A
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465A-1

Query Match 13.7%; Score 32.8; DB 17; Length 2731748;
Best Local Similarity 53.0%; Pred. No. 61;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy         40 CGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGA 99
              | ||  |   | ||||||  | |  ||||       | | | ||  || ||  || | |
Db     547279 CCTCACCGATGGCAACGGCTGGCCCGTTTCAATCATTAAAAAAGGCAATACAGACATTGT 547220

Qy        100 TAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGA 159
               ||||  || |||||    |  ||  |  |||||||  | |      || |||||||
Db     547219 CATCCCTAAACAGTCGCAGCGTCATCACAATCAGCAAACCACCACCCGGCTGTTTCAAGC 547160

Qy        160 ACTGTGTGGCAA 171
              |   ||  |||
Db     547159 AACCTGCACCAA 547148
```

RESULT 5
US-10-320-797-64/c
; Sequence 64, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 3250
; TYPE: DNA
; ORGANISM: Cryptococcus neoformans
US-10-320-797-64

Query Match 13.4%; Score 32; DB 16; Length 3250;
Best Local Similarity 52.2%; Pred. No. 4.2;
Matches 71; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy          7 CACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGT 66
              || | | | |  | | | |   |||||  ||  |||| | | |||| ||||    ||  |
Db        395 CAGGGCATCGGCATCCCCCTCACAGCAGATCTACGTCCTAACCGCCGACGGTTCACGCCT 336

Qy         67 TTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCT 126
                || |   ||||| |    | | |  |||  ||    |||| ||| |       | | |
Db        335 CTTCCTCCTCGATCCTACAAGACCGAACGAGGAAGAACCTCCACCATACGCATCTTTCCC 276

Qy        127 TAGAAAGCAAACAGCT 142
              || | | || |  ||
Db        275 TATACAACATATCCCT 260
```

RESULT 6
US-10-282-122A-24307
; Sequence 24307, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24307
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Listeria monocytogenes
US-10-282-122A-24307

Query Match 13.2%; Score 31.6; DB 13; Length 969;
Best Local Similarity 65.7%; Pred. No. 3.2;
Matches 46; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy        108 GACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTG 167
              || |   | | || ||| |||||||   |  | |||||| | || || || ||  ||
Db          7 GAACAAAGAGAAAAATTATTAGAAATTCATAATCTAAAGCAATACTTTAACAAAGGTACT 66

Qy        168 GCAAGTGAAG 177
              ||||||||||
Db         67 GCAAGTGAAG 76
```

RESULT 7
US-10-767-701-1632/c
; Sequence 1632, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 1632
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS27424_1
US-10-767-701-1632

Query Match 13.1%; Score 31.4; DB 17; Length 661;
Best Local Similarity 56.2%; Pred. No. 3.1;
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy         52 CAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACC 111
              ||  |  |  | | || |||  ||| || ||  | || || | |  | ||||| |  | |
Db        654 CAGAGCCAAACATTTTATTCCATTCCATTCAATTCAATCCTATGGAGCTAGCCATTCATC 595

Qy        112 GTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCA 156
              ||   |  ||   | ||| |   | |  |||||| || ||| ||
Db        594 ATCAGTTGCTACATTCAGATATGCACCTCCTAAAGCACAATTACA 550
```

RESULT 8
US-10-390-184-2
; Sequence 2, Application US/10390184
; Publication No. US20040053387A1
; GENERAL INFORMATION:
; APPLICANT: Kumiai Chemical Industry Co., Ltd.
; APPLICANT: Hayashi, Hiroaki
; TITLE OF INVENTION: A Sieve Tube-Localized Glutathione S-Transferase
; FILE REFERENCE: PH-1764 US
; CURRENT APPLICATION NUMBER: US/10/390,184
; CURRENT FILING DATE: 2003-03-18
; PRIOR APPLICATION NUMBER: JP 2002-268680
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1205
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-390-184-2

Query Match 13.1%; Score 31.4; DB 13; Length 1205;
Best Local Similarity 57.7%; Pred. No. 4.1;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy         34 CGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA 93
              |||||||||||||| ||   ||| |||  |||   ||   |||||   || ||| | ||
Db       1035 CGTCGTCGTCGTCGTCAATAACCGGACACGTACGTTCGTGTTCGCGCCATATATATCCTG 1094

Qy         94 CGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGA 130
```

```
Db       1095 CTGCTGTTGGTGAATTCCGAGTGGTCACATTATTATA 1131
```

RESULT 9
US-10-437-963-35457/c
; Sequence 35457, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 35457
; LENGTH: 1207
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_39375C.1
US-10-437-963-35457

Query Match 13.1%; Score 31.4; DB 17; Length 1207;
Best Local Similarity 57.7%; Pred. No. 4.1;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy          34 CGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA 93
               ||||||||||||| ||   ||| |||  |||   ||   |||||   || ||| | ||
Db         159 CGTCGTCGTCGTCGTCAATAACCGGACACGTACGTTCGTGTTCGCGCCATATATATCCTG 100

Qy          94 CGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGA 130
               |  |  | |       |||    || | ||| ||| |
Db          99 CTGCTGTTGGTGAATTCCGAGTGGTCACATTATTATA 63
```

RESULT 10
US-10-390-184-3
; Sequence 3, Application US/10390184
; Publication No. US20040053387A1
; GENERAL INFORMATION:
; APPLICANT: Kumiai Chemical Industry Co., Ltd.
; APPLICANT: Hayashi, Hiroaki
; TITLE OF INVENTION: A Sieve Tube-Localized Glutathione S-Transferase
; FILE REFERENCE: PH-1764 US
; CURRENT APPLICATION NUMBER: US/10/390,184
; CURRENT FILING DATE: 2003-03-18
; PRIOR APPLICATION NUMBER: JP 2002-268680
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1238
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-390-184-3

Query Match 13.1%; Score 31.4; DB 13; Length 1238;
Best Local Similarity 57.7%; Pred. No. 4.2;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy          34 CGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA 93
               ||||||||||||| ||   ||| |||  |||   ||   |||||   || ||| | ||
Db       1052 CGTCGTCGTCGTCGTCAATAACCGGACACGTACGTTCGTGTTCGCGCCATATATATCCTG 1111

Qy          94 CGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGA 130
               |  |  | |       |||    || | ||| ||| |
Db       1112 CTGCTGTTGGTGAATTCCGAGTGGTCACATTATTATA 1148
```

RESULT 11
US-10-175-523-86
; Sequence 86, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 86
; LENGTH: 158405
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-86

Query Match 13.1%; Score 31.4; DB 15; Length 158405;
Best Local Similarity 59.6%; Pred. No. 47;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy         128 AGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGT 187
               | ||| |||  |    || ||  |   ||| |||  || |||| | ||||||||||| ||
Db       27640 AAAAACCAACAACAACAAAAAACAAAACAACAACCCTGCGGCACGGGAAGGGTAGTGAGC 27699

Qy         188 GATACACCGGTTATATCGGAGTGGCGAGA 216
               | ||  || ||  ||  ||  ||  ||
Db       27700 TGTGCAGAGGCTAACTCAAAGCAGCTGGA 27728
```

RESULT 12
US-10-239-676-40/c
; Sequence 40, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06

```
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 40
;   LENGTH: 6061
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-40

  Query Match            13.0%;  Score 31;  DB 15;  Length 6061;
  Best Local Similarity   64.8%;  Pred. No. 13;
  Matches   46;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         34 CGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA 93
              |||||||| || ||||  | ||| || |||  | ||||| ||| |   | | ||| ||
Db       2164 CGTCGTCGCCGCCATCCTCGACGCGACTCGCTTCTTTCGATTCTACCTAAAAAAATCCAC 2105

Qy         94 CGTCGATAGCC 104
              || |   |||
Db       2104 CGCCATCCGCC 2094


RESULT 13
US-10-437-963-92822
; Sequence 92822, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 92822
;   LENGTH: 498
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_91267C.1
US-10-437-963-92822

  Query Match            12.6%;  Score 30.2;  DB 17;  Length 498;
  Best Local Similarity   60.2%;  Pred. No. 7;
  Matches   50;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         21 CGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATA 80
              ||||| |||   |||||||||||||||| ||| | |||  ||  | | |   |  || |
Db         94 CGCGCTGGCGCGACGTCGTCGTCGTCGTCGACGACGACAGCGGAGGTGACCACTTGGCAG 153

Qy         81 CAGATAAAACCCACGTCGATAGC 103
              |||   ||  |  || ||   ||
Db        154 CAGGCCAAGGCGGCGCCGGCGGC 176


RESULT 14
US-10-027-632-203363
; Sequence 203363, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203363
;   LENGTH: 616
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-203363

  Query Match            12.6%;  Score 30.2;  DB 13;  Length 616;
  Best Local Similarity   56.6%;  Pred. No. 7.8;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        111 CGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCA 170
              | | ||  ||  ||||||  |||| || ||    ||||   || | |  |||||||
Db        275 CTTTGCAGAAGGTTCTTATCAAGCTAAGAGGGGCAGAAAATTTCCTATTACTGTGTCATG 334

Qy        171 AGTGAAGGGTAGTTAGTGATACACCGGTTATATCGGAGT 209
              | || |||| |   |||  | ||  ||| | |  | |
Db        335 CATAAATGGTAATGCCTGAACCTCCACTTACAGCATATT 373


RESULT 15
US-10-027-632-203363
; Sequence 203363, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203363
;   LENGTH: 616
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-203363

  Query Match            12.6%;  Score 30.2;  DB 16;  Length 616;
  Best Local Similarity   56.6%;  Pred. No. 7.8;
```

Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```
Qy        111 CGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCA 170
              | | ||  ||  ||||||  |||| || ||    ||||   || | |  |||||||
Db        275 CTTTGCAGAAGGTTCTTATCAAGCTAAGAGGGGCAGAAAATTTCCTATTACTGTGTCATG 334

Qy        171 AGTGAAGGGTAGTTAGTGATACACCGGTTATATCGGAGT 209
              | || |||| |   |||  | ||  ||| | |  | |
Db        335 CATAAATGGTAATGCCTGAACCTCCACTTACAGCATATT 373
```

Search completed: September 18, 2004, 15:14:57
Job time : 269.741 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 10:33:16 ; Search time 2018.88 Seconds
(without alignments)
3535.155 Million cell updates/sec

Title: US-10-614-282-1
Perfect score: 239
Sequence: 1 atcagtcacgacttggtaag.....tgtggttccggctggacaat 239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	239	100.0	510	12	BI375946	BI375946 RE63854.5
c	2	37.2	15.6	1138	13	BQ144508	BQ144508 NF087E01D
	3	35.6	14.9	162	13	BW068571	BW068571 BW068571
	4	35	14.6	496	10	BF024030	BF024030 PvP 300 L
c	5	34.8	14.6	625	14	CD648340	CD648340 AUF_102_H
c	6	34.8	14.6	1201	13	BX420094	BX420094 BX420094
	7	34.6	14.5	563	28	AZ581975	AZ581975 1M0370K21
c	8	34	14.2	670	29	LBAF025H06	BX542182 Leishmani
c	9	34	14.2	670	29	LBAF025H09	BX542273 Leishmani
	10	33.8	14.1	507	29	BX137891	BX137891 Danio rer
	11	33.6	14.1	638	28	AZ983355	AZ983355 2M0264C01
c	12	33.4	14.0	1026	11	AF227817	AF227817 Ictalurus
	13	33.2	13.9	537	13	BQ743358	BQ743358 WHE4103_A
	14	32.8	13.7	349	13	BU578178	BU578178 sar48g12.
	15	32.8	13.7	438	10	AW133037	AW133037 se13e10.y
	16	32.8	13.7	443	10	BE803253	BE803253 sr54a06.y
	17	32.8	13.7	446	12	BG725287	BG725287 sae34h10.
	18	32.8	13.7	773	28	BH465834	BH465834 BOHJI55TR
	19	32.8	13.7	955	13	BU901940	BU901940 AGENCOURT
c	20	32.6	13.6	624	29	BX139003	BX139003 Danio rer
	21	32.6	13.6	852	14	CD649081	CD649081 AUF_104_I
	22	32.6	13.6	895	14	CD048532	CD048532 AGENCOURT
c	23	32.2	13.5	565	13	BY492451	BY492451 BY492451
c	24	32.2	13.5	648	29	CE691717	CE691717 tigr-gss-
c	25	32.2	13.5	664	10	BE941338	BE941338 EST420917
	26	32.2	13.5	1201	9	AL573127	AL573127 AL573127
	27	32	13.4	134	9	AU077131	AU077131 AU077131
	28	32	13.4	608	12	BI561029	BI561029 603254230
	29	32	13.4	726	10	BG037177	BG037177 602288557
	30	32	13.4	775	12	BG616128	BG616128 602642702
	31	32	13.4	786	13	BX116922	BX116922 BX116922
c	32	31.8	13.3	731	28	BH992982	BH992982 oed84e02.
c	33	31.8	13.3	739	13	BX605323	BX605323 BX605323
	34	31.8	13.3	776	29	CC532579	CC532579 CH240_409
c	35	31.8	13.3	1267	28	CC324651	CC324651 TAM32-13M
	36	31.6	13.2	354	9	AU273213	AU273213 AU273213
	37	31.6	13.2	653	28	AZ760934	AZ760934 1M0555E02
c	38	31.6	13.2	739	14	CA413042	CA413042 UI-H-EZ0-
	39	31.6	13.2	1077	13	BQ952602	BQ952602 AGENCOURT
	40	31.4	13.1	364	9	AU082434	AU082434 AU082434
	41	31.4	13.1	425	28	AZ404419	AZ404419 1M0172O23
	42	31.4	13.1	466	10	AW448888	AW448888 BRY_1810
c	43	31.4	13.1	549	10	BF657587	BF657587 OV2_17_A0
	44	31.4	13.1	555	9	AU082410	AU082410 AU082410
c	45	31.4	13.1	654	14	CD233195	CD233195 SS1_12_E0

ALIGNMENTS

RESULT 1
BI375946
LOCUS BI375946 510 bp mRNA linear EST 01-AUG-2001
DEFINITION RE63854.5prime RE Drosophila melanogaster normalized Embryo pFlc-1 Drosophila melanogaster cDNA clone RE63854 5, mRNA sequence.
ACCESSION BI375946
VERSION BI375946.1 GI:15071974
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 510)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
TITLE BDGP/HHMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798

```
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003674: arm:3R [2306679,2604706]
            estimated-cyto:83F1-84A5: 05/16/2001
            Plate: RE.638  row: E  column: 6
            High quality sequence stop: 453.
FEATURES              Location/Qualifiers
     source           1. .510
                      /organism="Drosophila melanogaster"
                      /mol_type="mRNA"
                      /db_xref="taxon:7227"
                      /clone="RE63854"
                      /sex="male and female"
                      /dev_stage="0-24 hours mixed stage embryonic"
                      /lab_host="DH5-alpha TonA"
                      /clone_lib="RE Drosophila melanogaster normalized Embryo
                      pFlc-1"
                      /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                      BamHI; Library was kindly generated by Piero Carninci at
                      the RIKEN. The library  was normalized and excised using
                      Cre recombinase. Plasmid cDNA library."
ORIGIN

  Query Match             100.0%;  Score 239;  DB 12;  Length 510;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-64;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 ATCAGTCACGACTTGGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAG 65

Qy         61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 TCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAA 125

Qy        121 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 TATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGT 185

Qy        181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 244


RESULT 2
BQ144508/c
LOCUS       BQ144508                1138 bp    mRNA    linear   EST 24-APR-2002
DEFINITION  NF087E01DT1F1005 Drought Medicago truncatula cDNA clone NF087E01DT
            5', mRNA sequence.
ACCESSION   BQ144508
VERSION     BQ144508.1  GI:20281567
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 1138)
  AUTHORS   Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula drought library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7391
            Fax: 580 221 7380
            Email: gdmay@noble.org
            Insert Length: 1138   Std Error: 0.00
            Plate: 087  row: E  column: 01
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES              Location/Qualifiers
     source           1. .1138
                      /organism="Medicago truncatula"
                      /mol_type="mRNA"
                      /db_xref="taxon:3880"
                      /clone="NF087E01DT"
                      /tissue_type="Plantlets"
                      /dev_stage="Pooled timepoints"
                      /clone_lib="Drought"
                      /note="Vector: Lambda Zap; Contains a mixture of entire
                      plantlets harvested in a series of days-post-watering
                      timepoints. "
ORIGIN

  Query Match             15.6%;  Score 37.2;  DB 13;  Length 1138;
  Best Local Similarity   50.6%;  Pred. No. 1.7;
  Matches   90;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy         15 GGTAAGCGCGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGT 74
              |||| ||| ||||||   | |||| ||||| |  |||   | | || ||  | ||    ||
Db        924 GGTACGCGGGCAGGCGTAATGTCGGCGTCGGCGACGCGCGCAGCAGACGGATATGAGCGT 865

Qy         75 TCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGC 134
              |||||   || | |  || |||||    | ||    | ||  ||     | ||  ||
Db        864 GCGATATTCATCATATAGACAGCGATATATATAGAGTCACACGCGATCAGGCTGGAGTGC 805

Qy        135 AAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGATAC 192
              |      | || |    | | || | | |  || ||||  |      | ||||
Db        804 GAGTTATGACAGGATAGATAGAGAAAGTATATACAAAATGAAACGATAGGGGGAATAC 747


RESULT 3
BW068571
LOCUS       BW068571                162 bp    mRNA    linear   EST 20-OCT-2002
DEFINITION  BW068571 Nori Satoh unpublished cDNA library, cleaving embryo Ciona
            intestinalis cDNA clone rciclo75j03 3', mRNA sequence.
ACCESSION   BW068571
VERSION     BW068571.1  GI:24169983
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 162)
  AUTHORS   Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona intestinalis (2002c)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .162
                      /organism="Ciona intestinalis"
                      /mol_type="mRNA"
                      /db_xref="taxon:7719"
                      /clone="rciclo75j03"
                      /tissue_type="whole body"
                      /dev_stage="cleaving embryo"
                      /clone_lib="Nori Satoh unpublished cDNA library, cleaving
                      embryo"
ORIGIN

  Query Match             14.9%;  Score 35.6;  DB 13;  Length 162;
  Best Local Similarity   50.0%;  Pred. No. 3;
  Matches   74;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         53 AACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCG 112
              || ||  | ||  |||          |    | ||||     |||| ||||      | |
Db          2 AATGGTNGACGCTTTTGCACAANNAANNAGGTNAAAAAGTTGGTCGGTAGCAAGGNAACT 61
```

```
Qy        113 TCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAG 172
              |  |  || ||    | ||||||||  |  ||   |  | | | ||  ||  |  ||||
Db         62 TTCCAAAAAATGNAAAAAAAGCAAAATNCCCAACCCCCTTGTNACGAGGTGGTTTACAAG 121

Qy        173 TGAAGGGTAGTTAGTGATACACCGGTTA 200
              ||      ||  || | ||||||||||
Db        122 AGAGCAACTGTGGGTTAAACACCGGTTA 149


RESULT 4
BF024030
LOCUS       BF024030                496 bp    mRNA    linear   EST 30-OCT-2000
DEFINITION  PvP 300 L99-29 Litopenaeus vannamei cDNA similar to 16S rRNA, mRNA
            sequence.
ACCESSION   BF024030
VERSION     BF024030.1  GI:11038174
KEYWORDS    EST.
SOURCE      Litopenaeus vannamei (Pacific white shrimp)
  ORGANISM  Litopenaeus vannamei
            Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
            Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
            Penaeidae; Litopenaeus.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Gross,P.S., Bartlett,T.C., Browdy,C.L., Chapman,R.W. and Warr,G.W.
  TITLE     Immune gene discovery by expressed sequence tag analysis of
            hemocytes and hepatopancreas in the Pacific White Shrimp,
            Litopenaeus vannamei, and the Atlantic White Shrimp, L. setiferus
  JOURNAL   Dev. Comp. Immunol. 25 (7), 565-577 (2001)
  MEDLINE   21365521
   PUBMED   11472779
COMMENT     Contact: Gross PS, Bartlett TC
            Department of Biochemistry and Molecular Biology
            Medical University of South Carolina
            Box 250509, 173 Ashley Avenue, Charleston, SC 29425, USA
            Tel: 846 792 9935
            Fax: 843 792 4850
            Email: grossp@musc.edu.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Litopenaeus vannamei"
                     /mol_type="mRNA"
                     /db_xref="taxon:6689"
                     /tissue_type="Hepatopancreas"
                     /clone_lib="L99-29"
                     /note="Vector: pTriplEx2; Site_1: Sfi I; Site_2: Sfi I"
ORIGIN

  Query Match            14.6%;  Score 35;  DB 10;  Length 496;
  Best Local Similarity  57.1%;  Pred. No. 6.4;
  Matches   56;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        115 GCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTG 174
              ||||||| ||||| || ||   ||| |   || |     |   || |    ||   | |
Db        248 GCGTAATCTTCTTTGAGAGTCCACATCGNCAGGAAGGGGTTGCGACCNCGATGNTGAATN 307

Qy        175 AAGGGTAGTTAGTGATACACCGGTTATATCGGAGTGGC 212
              ||||||   |  |  | |||| | ||||  |||| |||
Db        308 AAGGGTTCCTTATANTGCACCNGNTATAAAGGAGGGGC 345


RESULT 5
CD648340/c
LOCUS       CD648340                625 bp    mRNA    linear   EST 18-JUN-2003
DEFINITION  AUF_102_H24_T7 Crassostrea virginica Gill Crassostrea virginica
            cDNA 5' similar to Unknown, mRNA sequence.
ACCESSION   CD648340
VERSION     CD648340.1  GI:31904281
KEYWORDS    EST.
SOURCE      Crassostrea virginica (eastern oyster)
  ORGANISM  Crassostrea virginica
            Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Ostreoida;
            Ostreoidea; Ostreidae; Crassostrea.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Peatman,E., Kucuktas,H., Li,P., He,C., Feng,J., Wei,X. and Liu,Z.
  TITLE     Differentially expressed oyster (Crassostrea virginica) genes after
            exposure to mercury
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Liu ZJ
            The Fish Molecular Genetics and Biotechnology Laboratory,
            Department of Fisheries and Allied Aquacultures and Program of Cell
            and Molecular Biosciences
            Auburn University
            203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
            Tel: 334 844 4054
            Fax: 334 844 9208
            Email: zliu@acesag.auburn.edu
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Crassostrea virginica"
                     /mol_type="mRNA"
                     /db_xref="taxon:6565"
                     /clone_lib="Crassostrea virginica Gill"
                     /note="Organ: Gill; Vector: pSport1; Site_1: NotI; Site_2:
                     SalI"
ORIGIN

  Query Match            14.6%;  Score 34.8;  DB 14;  Length 625;
  Best Local Similarity  57.3%;  Pred. No. 7.9;
  Matches   63;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        128 AGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGT 187
              | |||||||| |   ||| ||||||||  | |          ||| || ||  | |||| |
Db        345 AAAAAGCAAATAAGAAAAAAACTATTTACACACACACAACACAATTGTAGCTTGGTTACT 286

Qy        188 GATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACA 237
              | | ||    | | |  ||    ||| | | ||  ||||| ||  ||||
Db        285 AAGAAACTACATTTCTTTGAAGTTCGATATACTGATGTTCCTGCGTGACA 236


RESULT 6
BX420094/c
LOCUS       BX420094                1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX420094 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF023YF10 5-PRIME, mRNA sequence.
ACCESSION   BX420094
VERSION     BX420094.1  GI:30638109
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2290.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF023DC05QP1&cluster=2290.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF023DC05QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF023YF10"
                     /tissue_type="FETAL BRAIN"
```

/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 14.6%; Score 34.8; DB 13; Length 1201;
Best Local Similarity 24.5%; Pred. No. 9.6;
Matches 38; Conservative 63; Mismatches 54; Indels 0; Gaps 0;

```
Qy          85 TAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAA 144
               :||||||:|:::  |:| :  : :| : :|: ::  :| :::: ::||     |   : |
Db         958 KAAAACCBAMKWAAAKACBAAKKKAABCBCBAKKKKKAABBBKCKDAAAAKKKCBAKKCA 899

Qy         145 AGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGATACACCGGTTATATC 204
               | || :|::  :  || :::::::: || :  |:::  :::|:  |  ::|: :: | |:
Db         898 AAAAADAKKAARKAAAADKKKKBSAAACBANAKBKAKKKDADAAAACBMCBAKDAAAAWK 839

Qy         205 GGAGTGGCGAGAAAGTGTGGTTCCGGCTGGACAAT 239
               : |:: :|:|:|  :   : ::::|  :: :| ||:
Db         838 KAAKKAKCDADACWAACKAKKKBCYYYKCBAAAAK 804
```

RESULT 7
AZ581975
LOCUS AZ581975 563 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0370K21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0370K21 R, genomic survey sequence.
ACCESSION AZ581975
VERSION AZ581975.1 GI:11697356
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 563)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0370 row: K column: 21
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 563.
FEATURES Location/Qualifiers
source 1. .563
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0370K21"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 14.5%; Score 34.6; DB 28; Length 563;
Best Local Similarity 51.3%; Pred. No. 8.9;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy          83 GATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCT 142
               | | || | ||   | ||  ||| | || |  | | | ||  ||| ||| || ||   |
Db         286 GCTTAACCACAGAGCCATCTCCCCAGGCCTCCTGGAAGTAAACTTTTAAACCACACTTTT 345

Qy         143 AAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGATACACCGGTTATA 202
               | ||  || ||| || |  ||| |  ||| |  |||   ||  | ||  ||  ||
Db         346 ATAGTTCTTTTTAAAAATATGTATTTCAATTATGTGGTGAGGAGGACTGCAAATGTCCTA 405

Qy         203 TCGGAGTGGCGAGAAAGTGTGGTTCCGGCTGGAC 236
                  |   | ||| | |||       ||||| |
Db         406 NAAGCTAGAAGAGGATGTGACCCCGGGGCTGGTC 439
```

RESULT 8
LBAF025H06/c
LOCUS LBAF025H06 670 bp DNA linear GSS 20-JUN-2003
DEFINITION Leishmania braziliensis GSS, clone LBAF025H06, genomic survey sequence.
ACCESSION BX542182
VERSION BX542182.1 GI:32136084
KEYWORDS GSS; genomic survey sequence.
SOURCE Leishmania braziliensis
ORGANISM Leishmania braziliensis
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania; Leishmania braziliensis species complex.
REFERENCE 1
AUTHORS Laurentino,E.C., Ruiz,J.C. and Cruz,A.K.
TITLE GSS analysis of the Leishmania braziliensis genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 670)
AUTHORS Cruz,A.K.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo, Department of Molecular and Cell Biology, FMRP, Avenida Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
COMMENT Clone requests: akcruz@fmrp.usp.br.
FEATURES Location/Qualifiers
source 1. .670
/organism="Leishmania braziliensis"
/mol_type="genomic DNA"
/strain="MHOM/BR/75/M2904"
/db_xref="taxon:5660"
/clone="LBAF025H06"

ORIGIN

Query Match 14.2%; Score 34; DB 29; Length 670;
Best Local Similarity 52.1%; Pred. No. 14;
Matches 76; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy          25 CAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGA 84
               ||  || | || |||||||  | ||| | ||   | |||| |    || | |   | |
Db         270 CATACACCGCGACGTCGTCCCCCTCGGCCACAACATACGTGGTGCGTGGCTTGCCGCGTA 211
```

```
Qy          85 TAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAA 144
               | | ||  |||| | ||  | | | |  | |||| | | |  || | |
Db         210 CCCAGCTCAGATCGAGTACGCTTCCCAGGCCCACCAGGGTCTTCGGTACCTGCGAGGTCA 151

Qy         145 AGAACTATTTCAAGAACTGTGTGGCA 170
               | | || || | ||| ||| |||
Db         150 CCCAGTCTTCCATGCACTTTGTAGCA 125
```

RESULT 9
LBAF025H09/c
LOCUS LBAF025H09 670 bp DNA linear GSS 20-JUN-2003
DEFINITION Leishmania braziliensis GSS, clone LBAF025H09, genomic survey sequence.
ACCESSION BX542273
VERSION BX542273.1 GI:32136112
KEYWORDS GSS; genomic survey sequence.
SOURCE Leishmania braziliensis
ORGANISM Leishmania braziliensis
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania; Leishmania braziliensis species complex.
REFERENCE 1
AUTHORS Laurentino,E.C., Ruiz,J.C. and Cruz,A.K.
TITLE GSS analysis of the Leishmania braziliensis genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 670)
AUTHORS Cruz,A.K.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo, Department of Molecular and Cell Biology, FMRP, Avenida Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
COMMENT Clone requests: akcruz@fmrp.usp.br.
FEATURES Location/Qualifiers
source 1. .670
/organism="Leishmania braziliensis"
/mol_type="genomic DNA"
/strain="MHOM/BR/75/M2904"
/db_xref="taxon:5660"
/clone="LBAF025H09"
ORIGIN

Query Match 14.2%; Score 34; DB 29; Length 670;
Best Local Similarity 52.1%; Pred. No. 14;
Matches 76; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy          25 CAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACAGA 84
               ||  || | || |||||||  | ||| | ||   | |||| |    || | |   | |
Db         270 CATACACCGCGACGTCGTCCCCCTCGGCCACAACATACGTGGTGCGTGGCTTGCCGCGTA 211

Qy          85 TAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAA 144
               | | ||  ||||  | ||  | | | |  |  |||| | | |  || | |
Db         210 CCCAGCTCAGATCGAGTACGCTTCCCAGGCCCACCAGGGTCTTCGGTACCTGCGAGGTCA 151

Qy         145 AGAACTATTTCAAGAACTGTGTGGCA 170
               | | || || | ||| ||| |||
Db         150 CCCAGTCTTCCATGCACTTTGTAGCA 125
```

RESULT 10
BX137891
LOCUS BX137891 507 bp DNA linear GSS 13-MAR-2003
DEFINITION Danio rerio genomic clone DKEY-86F3, genomic survey sequence.
ACCESSION BX137891
VERSION BX137891.1 GI:27969207
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 507)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 86F3. 86F3 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES Location/Qualifiers
source 1. .507
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-86F3"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"
ORIGIN

Query Match 14.1%; Score 33.8; DB 29; Length 507;
Best Local Similarity 51.0%; Pred. No. 15;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy          53 AACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCG 112
               || | | |  ||  ||||          ||  |||    ||  |  ||||  |||    |
Db         336 AATGCGTGCAGTCATTTTGACCGTTGGTCATTTAAGGGACAGATACATAGAACTCTTTTG 395

Qy         113 TCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAG 172
               |    || |||| | | |||   ||||     ||||  | ||  | || | | ||| ||
Db         396 TGTTTTATTATTTTAATAAAAATAACAATGTTAGAATAAATTATACAAATATTTGGGAAA 455

Qy         173 TGAAGGGTAGTTAGTGATACACCGGTTATATCGGAGT 209
               |   |  | |||  | || |  |||| |||  || |
Db         456 AGTCAGTGAATTATAGCTATATGGGTTTTATTTGAAT 492
```

RESULT 11
AZ983355
LOCUS AZ983355 638 bp DNA linear GSS 27-APR-2001
DEFINITION 2M0264C01R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0264C01 R, genomic survey sequence.
ACCESSION AZ983355
VERSION AZ983355.1 GI:13854582
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 638)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0264 row: C column: 01
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 638.
FEATURES Location/Qualifiers
source 1. .638
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"

/db_xref="taxon:10090"
/clone="UUGC2M0264C01"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 14.1%; Score 33.6; DB 28; Length 638;
Best Local Similarity 61.4%; Pred. No. 19;
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy        130 AAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGTTAGTGA 189
              ||||||  |   | ||| | |||||  ||  ||||||  || | | |||  | |||||  |
Db        310 AAAGCACAGTTCCAAAAATCTATTAAAACTACTGTGGTGCTAATAAAGAATTGTTAGATA 369

Qy        190 TACACCGGTTATATCGGAGTGGCGAGAA 217
              ||  | |  ||| | |||    | | ||
Db        370 TATTCTGTGTATTTTGGAAAACCTATAA 397
```

RESULT 12
AF227817/c
LOCUS AF227817 1026 bp mRNA linear HTC 23-JAN-2003
DEFINITION Ictalurus punctatus mitochondrial long-chain enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase mRNA, partial sequence.
ACCESSION AF227817
VERSION AF227817.1 GI:27883582
KEYWORDS HTC.
SOURCE Ictalurus punctatus (channel catfish)
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes; Ictaluridae; Ictalurus.
REFERENCE 1 (bases 1 to 1026)
AUTHORS Kim,S., Li,P., Zheng,X., Dunham,R.A. and Liu,Z.
TITLE Gene expression in the muscles of young and mature channel catfish (Ictalurus punctatus) as analyzed by expressed sequence tags and gene filters
JOURNAL Fish Physiol. Biochem. (2003) In press
REFERENCE 2 (bases 1 to 1026)
AUTHORS Liu,Z. and Kim,S.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2000) Fisheries and Allied Aquacultures and the Program of Cell and Molecular Biosciences, Auburn University, 203 Swingle Hall, Auburn, AL 36849, USA
FEATURES Location/Qualifiers
source 1. .1026
/organism="Ictalurus punctatus"
/mol_type="mRNA"
/db_xref="taxon:7998"
/tissue_type="muscle"
misc_feature <1. .>1026
/note="similar to mitochondrial long-chain enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; coding region not determined"

ORIGIN

Query Match 14.0%; Score 33.4; DB 11; Length 1026;
Best Local Similarity 65.3%; Pred. No. 25;
Matches 49; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
Qy        110 CCGTCGCGTAATATTCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGC 169
              || |  |||  | |||||||   || | | ||   |||| |||| | |||||||| |
Db        911 CCCTTCTGTAGAACTCTTAGATCACACAGACCTTCGGAACGATTTTAGGAACTGTGTAGT 852

Qy        170 AAGTGAAGGGTAGTT 184
              ||| |||  || | |
Db        851 AAGAGAAATGTCGGT 837
```

RESULT 13
BQ743358
LOCUS BQ743358 537 bp mRNA linear EST 17-JUL-2002
DEFINITION WHE4103_A10_B19ZS Wheat salt-stressed root cDNA library Triticum aestivum cDNA clone WHE4103_A10_B19, mRNA sequence.
ACCESSION BQ743358
VERSION BQ743358.1 GI:21890145
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 537)
AUTHORS Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J., Lazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
TITLE The structure and function of the expressed portion of the wheat genomes - Salt-stressed root cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES Location/Qualifiers
source 1. .537
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4103_A10_B19"
/tissue_type="Roots"
/dev_stage="Full tillering"
/lab_host="E. coli SOLR"
/clone_lib="Wheat salt-stressed root cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Hydroponic plants grown to full tillering stage were treated with 150 mM NaCl for either 12 hours or 7 days. Root tissues of the plants subjected to both types of treatment were collected separately at University of California, Davis (E. Akhunov and K. Deal in J. Dvorak's Lab). Total RNA was prepared separately from the two samples (12h and 7day treatments), and equal amount of RNA was then pooled. PolyA RNA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak) at the University of California, Davis. Colony plating, plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

```
  Query Match           13.9%;  Score 33.2;  DB 13;  Length 537;
  Best Local Similarity  59.6%;  Pred. No. 24;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         23 CGCAGGCAGCACGTCGTCGTCGTCATCGCCAACGGGAGTCGTGTTTTTCGGTTCGATACA 82
              ||| || |||||||||||||||| |||||||  |||  ||| ||   | ||     || |
Db         74 CGCCGGAAGCACGTCGTCGTCGCCATCGCCGCCGGTGGTCCTGCCGTCCGAACAGAAAGC 133

Qy         83 GATAAAACCCACGTCGATAGCCCTCGACCGTCGC 116
                 |   || |  |   | || ||  ||| ||
Db        134 ACGGTACAGCAGGGAGGAGGACCCCGGTCGTTGC 167


RESULT 14
BU578178
LOCUS       BU578178                 349 bp    mRNA    linear   EST 17-SEP-2002
DEFINITION  sar48g12.y1 Gm-c1074 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1074-5711 5' similar to TR:O80921 O80921 PUTATIVE RNA-BINDING
            PROTEIN. ;, mRNA sequence.
ACCESSION   BU578178
VERSION     BU578178.1  GI:23057504
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 295.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1074-5711"
                     /tissue_type="seedlings induced for HR (hypersensitive
                     response)"
                     /dev_stage="9-11 day old"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1074"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from 9-11 day old seedlings that were induced for HR
                     (hypersensitive response) by vacuum infiltrating plant
                     tissue with Pseudomonas syringae pv. glycinea carrying the
                     avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
                     unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
                     and 53 hrs after inoculation and their mRNA pooled equally
                     for cDNA construction. The  library was prepared using the
                     Stratagene pBluescript II SK(+) library construction kit.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly(dT) sequence with an XhoI restriction
                     site. EcoRI adaptors were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA insert is
                     protected from XhoI digestion via methylation during first
                     strand synthesis. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into E.coli ElectroMax DH10B host cells. Plant
                     care, inoculations, and library construction were
                     performed by Steve Clough (Lila Vodkin lab, University of
                     Illinois)."
ORIGIN

  Query Match           13.7%;  Score 32.8;  DB 13;  Length 349;
  Best Local Similarity  61.9%;  Pred. No. 28;
  Matches   52;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        124 TCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGT 183
              |||||| || |  ||  ||| ||||    ||| |||   || ||  ||||   | | |||
Db        105 TCTTAGCAAACGTACTACTACAGAAAGGCTTCGAGAGGAGTTTGCAAAGTTTGGTGAAGT 164

Qy        184 TAGTGATACACCGGTTATATCGGA 207
              |  | || ||  |||| || | ||
Db        165 TGTTCATGCAAGGGTTGTAACTGA 188


RESULT 15
AW133037
LOCUS       AW133037                 438 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  se13e10.y1 Gm-c1013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1013-3427 5' similar to TR:Q9ZUR6 Q9ZUR6 PUTATIVE RNA BINDING
            PROTEIN. ;, mRNA sequence.
ACCESSION   AW133037
VERSION     AW133037.1  GI:6134644
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 772   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 271.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1013-3427"
                     /tissue_type="Whole seedlings, 2-3 week old seedlings,
                     greenhouse grown"
                     /lab_host="XL10-Gold"
                     /clone_lib="Gm-c1013"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
```

XhoI; This cDNA library was constructed from mRNA isolated from whole seedlings of 2-3 week old greenhouse grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

ORIGIN

Query Match 13.7%; Score 32.8; DB 10; Length 438;
Best Local Similarity 61.9%; Pred. No. 30;
Matches 52; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy        124 TCTTAGAAAGCAAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGGTAGT 183
              |||||| || |  ||  ||| ||||   ||| |||  || ||  ||||   | | |||
Db        213 TCTTAGCAAACGTACTACTACAGAAAGGCTTCGAGAGGAGTTTGCAAAGTTTGGTGAAGT 272

Qy        184 TAGTGATACACCGGTTATATCGGA 207
              |  | || ||  |||| || | ||
Db        273 TGTTCATGCAAGGGTTGTAACTGA 296
```

Search completed: September 18, 2004, 13:16:48
Job time : 2023.88 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 09:46:06 ; Search time 1166.45 Seconds
(without alignments)
9066.559 Million cell updates/sec

Title: US-10-614-282-2
Perfect score: 244
Sequence: 1 agcgccggggatttaaatgc.....accaggtcactcagtgacag 244

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	244	100.0	102717	9	AC004079	AC004079 Homo sapi
c 2	242.4	99.3	170237	9	AC142285	AC142285 Pan trogl
c 3	240.8	98.7	191971	9	AC146102	AC146102 Pan trogl
4	204.8	83.9	128851	2	AC129963	AC129963 Sus scrof
5	199.2	81.6	112570	10	AC116662	AC116662 Rattus no
6	199.2	81.6	234603	2	AC097087	AC097087 Rattus no
7	199.2	81.6	298853	2	AC122669	AC122669 Rattus no
8	184	75.4	200633	2	AC015583	AC015583 Mus muscu
9	184	75.4	236214	2	AC091106	AC091106 Mus muscu
10	184	75.4	236829	2	AC123672	AC123672 Mus muscu
c 11	172.6	70.7	6118	6	AX344703	AX344703 Sequence
12	135.2	55.4	6118	6	AX344702	AX344702 Sequence
13	62	25.4	2529	9	BC032547	BC032547 Homo sapi
14	53.8	22.0	2216	10	MUSERA1A	M22115 Mouse ERA-1
15	51	20.9	443	10	MUSHOX161	M20214 Mus musculu
c 16	44	18.0	209926	2	AC145789	AC145789 Silurana
c 17	41.4	17.0	160516	10	AL928943	AL928943 Mouse DNA
18	40.8	16.7	204697	2	AC120995	AC120995 Rattus no
c 19	40.8	16.7	236070	2	AC112571	AC112571 Rattus no
20	38.8	15.9	158174	8	AP003414	AP003414 Oryza sat
21	38.8	15.9	159761	8	AP003921	AP003921 Oryza sat
c 22	38	15.6	41722	9	AL359316	AL359316 Human DNA
c 23	38	15.6	302250	6	AX335531	AX335531 Sequence
c 24	38	15.6	324604	9	U82671	U82671 Homo sapien
25	37.8	15.5	77457	1	AF210249	AF210249 Streptomy
c 26	37.6	15.4	244549	2	AC147476	AC147476 Mus muscu
27	37.4	15.3	215259	2	AC118349	AC118349 Rattus no
28	37.2	15.2	719	11	PM7B12B	AL685078 Penicilli
29	37.2	15.2	3050	6	AX746737	AX746737 Sequence
30	37.2	15.2	3050	9	AK091024	AK091024 Homo sapi
31	37.2	15.2	174367	10	AL512597	AL512597 Mouse DNA
32	37	15.2	16115	1	AE007075	AE007075 Mycobacte
33	37	15.2	63155	6	AX646257	AX646257 Sequence
34	37	15.2	63155	9	AB065664	AB065664 Homo sapi
c 35	37	15.2	172953	9	AC067721	AC067721 Homo sapi
36	37	15.2	306050	1	BX248341	BX248341 Mycobacte
37	37	15.2	348247	15	BX842579	Bx842579 Mycobacte
38	36.8	15.1	1023	10	MMHOX16R	X06024 Mouse Hox-1
c 39	36.6	15.0	172945	9	AC007220	AC007220 Homo sapi
c 40	36.6	15.0	185908	9	AC136297	AC136297 Homo sapi
41	36.6	15.0	206894	10	AC109152	AC109152 Mus muscu
c 42	36.6	15.0	253850	2	AC096825	AC096825 Rattus no
c 43	36.6	15.0	257407	2	AC109870	AC109870 Rattus no
44	36.6	15.0	258319	2	AC127770	AC127770 Rattus no
c 45	36.6	15.0	266342	2	AC109963	AC109963 Rattus no

ALIGNMENTS

RESULT 1
AC004079/c
LOCUS AC004079 102717 bp DNA linear PRI 26-APR-2003
DEFINITION Homo sapiens PAC clone RP1-167F23 from 7, complete sequence.
ACCESSION AC004079
VERSION AC004079.1 GI:2822174
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 102717)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

```
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 102717)
  AUTHORS   Duckels,G., Hawkins,M., Hinds,K. and Jones,K.
  TITLE     The sequence of Homo sapiens PAC clone RP1-167F23
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
              Center project name: H_DJ0167F23
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone was provided for sequencing by Dr. Stephen Scherer,
            Department of Genetics, The Hospital for Sick Children, Toronto,
            Ontario, Canada, with support from the Canadian Genome Analysis and
            Technology Program; and Dr. John D. McPherson, Department of
            Genetics, Washington University, St. Louis MO. For additional
            information about the map position of this sequence, see
            http://www.genet.sickkids.on.ca/chromosome7 and
            http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at http://www.chori.org using the
            method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
            The library is from one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of RP1-167F23
            the actual end is at base position 102717 of RP1-167F23. The
            orientation of this clone is unknown.
FEATURES              Location/Qualifiers
     source           1. .102717
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /map="7"
                      /clone="RP1-167F23"
                      /clone_lib="RPCI-1"
     repeat_region    1. .1098
                      /rpt_family="L1"
     repeat_region    1099. .1412
                      /rpt_family="Alu"
     repeat_region    1413. .2186
                      /rpt_family="L1"
     repeat_region    2187. .2495
                      /rpt_family="Alu"
     repeat_region    2496. .2910
                      /rpt_family="L1"
     repeat_region    3131. .3332
                      /rpt_family="Alu"
     variation        3255
                      /allele="G"
                      /allele="A"
                      /db_xref="dbSNP:2428420"
     repeat_region    3333. .3696
                      /rpt_family="MaLR"
     variation        3645
                      /allele="C"
                      /allele="T"
                      /db_xref="dbSNP:2522823"
     repeat_region    3697. .3799
                      /rpt_family="Alu"
     repeat_region    4056. .4081
                      /rpt_family="(TTTC)n"
     repeat_region    4145. .4277
                      /rpt_family="L2"
     variation        4569
                      /allele="A"
                      /allele="C"
                      /db_xref="dbSNP:2522824"
     repeat_region    4926. .5064
                      /rpt_family="MER1_type"
     repeat_region    5081. .5332
                      /rpt_family="L1"
     variation        5625
                      /allele="T"
                      /allele="C"
                      /db_xref="dbSNP:2462901"
     repeat_region    6427. .6546
                      /rpt_family="Alu"
     repeat_region    6572. .6868
                      /rpt_family="Alu"
     variation        6927
                      /allele="A"
                      /allele="G"
                      /db_xref="dbSNP:1051188"
     variation        6969
                      /allele="G"
                      /allele="A"
                      /db_xref="dbSNP:1063644"
     variation        7372
                      /allele="T"
                      /allele="C"
                      /db_xref="dbSNP:2428430"
     variation        7487
                      /allele="A"
                      /allele="T"
                      /db_xref="dbSNP:1051239"
     variation        7510
                      /allele="C"
                      /allele="A"
                      /db_xref="dbSNP:1051241"
     repeat_region    7942. .8236
```

```
                     /rpt_family="Alu"
     repeat_region   8544. .8579
                     /rpt_family="AT_rich"
     variation       8760
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:2465263"
     repeat_region   8991. .9300
                     /rpt_family="Alu"
     variation       9745
                     /allele="G"
                     /allele="C"
                     /db_xref="dbSNP:2522829"
     repeat_region   10140. .10406
                     /rpt_family="L1"
     repeat_region   10407. .10702
                     /rpt_family="Alu"
     variation       10602
                     /allele="G"
                     /allele="A"
                     /db_xref="dbSNP:2462902"
     variation       10666
                     /allele="G"
                     /allele="A"
                     /db_xref="dbSNP:2254463"
     repeat_region   10703. .10790
                     /rpt_family="L1"
     repeat_region   11252. .11503
                     /rpt_family="MIR"
     variation       11267
                     /allele="C"
                     /allele="G"
                     /db_xref="dbSNP:2522822"
     repeat_region   11814. .11854
                     /rpt_family="MIR"
     repeat_region   12097. .12443
                     /rpt_family="Alu"
     variation       12465
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:2428416"

  Query Match             100.0%;  Score 244;  DB 9;  Length 102717;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-60;
  Matches  244;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78551 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 78492

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78491 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 78432

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78431 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 78372

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78371 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 78312

Qy        241 ACAG 244
              ||||
Db      78311 ACAG 78308


RESULT 2
AC142285/c
LOCUS       AC142285              170237 bp    DNA     linear   PRI 26-SEP-2003
DEFINITION  Pan troglodytes BAC clone RP43-169B4 from 7, complete sequence.
ACCESSION   AC142285
VERSION     AC142285.1  GI:29294404
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 170237)
  AUTHORS   Belter,E. and Maupin,R.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-169B4
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 170237)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Sequencing of Pan troglodytes
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 170237)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 170237)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT169B04
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.

            Discrepant bases between whole genome shotgun data and clone
            sequence.

            Data from RP43-21L13 was used to finish this clone.
FEATURES              Location/Qualifiers
     source           1. .170237
                      /organism="Pan troglodytes"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9598"
                      /chromosome="7"
                      /map="7"
```

```
                     /clone="RP43-169B4"
                     /clone_lib="RPCI-43"
ORIGIN

  Query Match             99.3%;  Score 242.4;  DB 9;  Length 170237;
  Best Local Similarity   99.6%;  Pred. No. 6.3e-60;
  Matches  243;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     119853 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 119794

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119793 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 119734

Qy        121 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119733 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 119674

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119673 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 119614

Qy        241 ACAG 244
              ||||
Db     119613 ACAG 119610


RESULT 3
AC146102/c
LOCUS       AC146102               191971 bp    DNA     linear   PRI 18-DEC-2003
DEFINITION  Pan troglodytes BAC clone RP43-21L13 from 7, complete sequence.
ACCESSION   AC146102
VERSION     AC146102.2  GI:38230057
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 191971)
  AUTHORS   Harkins,R. and Haakenson,W.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-21L13
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 191971)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Pan troglodytes
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 191971)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 191971)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 191971)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 9, 2003 this sequence version replaced gi:33387055.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT021L13
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .191971
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-21L13"
                     /clone_lib="RPCI-43"
     misc_feature    82205. .82239
                     /note="Sequence derived from PCR product of project DNA."
     unsure          156808. .156822
                     /note="Sequence derived from one plasmid subclone."
     unsure          157072. .157109
                     /note="Sequence derived from one plasmid subclone."
     unsure          159887. .160183
                     /note="Sequence derived from one plasmid subclone."
     unsure          177866. .177987
                     /note="Sequence derived from one plasmid subclone."
     unsure          178716. .178743
                     /note="Sequence derived from one plasmid subclone."
ORIGIN

  Query Match             98.7%;  Score 240.8;  DB 9;  Length 191971;
  Best Local Similarity   99.2%;  Pred. No. 1.9e-59;
  Matches  242;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      99935 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 99876

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99875 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 99816

Qy        121 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      99815 CCCTCCCGCCCCCCCTTTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 99756

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99755 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 99696

Qy        241 ACAG 244
```

```
                ||||
Db      99695 ACAG 99692


RESULT 4
AC129963
LOCUS       AC129963               128851 bp    DNA     linear   HTG 16-AUG-2002
DEFINITION  Sus scrofa clone RP44-485M13, *** SEQUENCING IN PROGRESS ***, 4
            unordered pieces.
ACCESSION   AC129963
VERSION     AC129963.2  GI:22267860
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 128851)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Sus scrofa, clone RP44-485M13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128851)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 128851)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 16, 2002 this sequence version replaced gi:22123285.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24464
                Center clone name: 485_M_13
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6913: contig of 6913 bp in length
            *     6914     7013: gap of 100 bp
            *     7014    10891: contig of 3878 bp in length
            *    10892    10991: gap of 100 bp
            *    10992    71101: contig of 60110 bp in length
            *    71102    71201: gap of 100 bp
            *    71202   128851: contig of 57650 bp in length.
FEATURES             Location/Qualifiers
     source          1. .128851
                     /organism="Sus scrofa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9823"
                     /clone="RP44-485M13"
                     /clone_lib="RPCI-44 Male Porcine BAC"
ORIGIN

  Query Match             83.9%;  Score 204.8;  DB 2;  Length 128851;
  Best Local Similarity   92.2%;  Pred. No. 6.1e-49;
  Matches  226;  Conservative    0;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||||||||||||||||||||||||| ||| |||||||||||||  || | |||| |||||||
Db     115286 AGCGCCGGGGATTTAAATGCCACTGAAAGGGTGATCCATCACCGCAGGAGCCAGCAAACT 115345

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGAC-ACCGTCACGTGCCCCTCCTCCAGCGTCCTCC 119
              || |||||||||| ||||||||||||  | |||||||||||||| | |||| |||||
Db     115346 TTCGCAGGAGGCTCCGCCATTGGCTGANGAGAGTCACGTGCCCCTCCCCTAGCGCCCTCC 115405

Qy        120 GCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115406 GCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCA 115465

Qy        180 TGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGT 239
              |||||||||||||||||||||||| ||||| | |||||||||||||| ||||||||||
Db     115466 TGGAGGAAGTGAGAAAGTTGGCACGGTCACTCAGGGCTTCGCAGGACCCGGTCACTCAGT 115525

Qy        240 GACAG 244
              |||||
Db     115526 GACAG 115530


RESULT 5
AC116662
LOCUS       AC116662               112570 bp    DNA     linear   ROD 02-MAY-2002
DEFINITION  Rattus norvegicus, clone RP31-256A2, complete sequence.
ACCESSION   AC116662
VERSION     AC116662.1  GI:19852147
KEYWORDS    HTG.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 112570)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Rattus norvegicus, clone RP31-256A2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112570)
  AUTHORS   Dewar,K., Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A.,
            Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J.,
```

```
            Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 112570)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3184
                Center clone name: 256_A_2
            --------------.
FEATURES             Location/Qualifiers
     source          1. .112570
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="RP31-256A2"
                     /clone_lib="RPCI-31 Female Rat PAC"
     repeat_region   480. .527
                     /rpt_family="(TG)n"
     repeat_region   2529. .2566
                     /rpt_family="AT_rich"
     repeat_region   4028. .4050
                     /rpt_family="GC_rich"
     repeat_region   8087. .8152
                     /rpt_family="CT-rich"
     repeat_region   10372. .10451
                     /rpt_family="(CCG)n"
     repeat_region   12554. .12595
                     /rpt_family="(TTTG)n"
     repeat_region   13284. .13307
                     /rpt_family="(TTTG)n"
     unsure          14249. .14319
                     /note="Less than 30 qual SNGL region"
     unsure          complement(14249. .14271)
                     /note="<30 quality SNGL region"
     unsure          complement(14278. .14288)
                     /note="<30 quality SNGL region"
     unsure          complement(14308. .14315)
                     /note="<30 quality SNGL region"
     repeat_region   15321. .15389
                     /rpt_family="CT-rich"
     repeat_region   15491. .15534
                     /rpt_family="(TTTC)n"
     repeat_region   16602. .16634
                     /rpt_family="AT_rich"
     repeat_region   16639. .16690
                     /rpt_family="(ATG)n"
     repeat_region   17526. .17595
                     /rpt_family="GA-rich"
     repeat_region   18502. .18556
                     /rpt_family="(TG)n"
     unsure          19757. .19765
                     /note="<30 quality SNGL region"
     repeat_region   20153. .20204
                     /rpt_family="(GA)n"
     repeat_region   20250. .20369
                     /rpt_family="(GAAA)n"
     repeat_region   21159. .21182
                     /rpt_family="AT_rich"
     repeat_region   23888. .23957
                     /rpt_family="(GA)n"
     repeat_region   27479. .27500
                     /rpt_family="GC_rich"
     repeat_region   28430. .28560
                     /rpt_family="CT-rich"
     repeat_region   29380. .29424
                     /rpt_family="(GGA)n"
     repeat_region   29526. .29558
                     /rpt_family="A-rich"
     repeat_region   30184. .30240
                     /rpt_family="(T)n"
     repeat_region   36278. .36303
                     /rpt_family="GC_rich"
     repeat_region   37035. .37151
                     /rpt_family="L1MDa"
     repeat_region   37173. .37238
                     /rpt_family="(GGAA)n"
     repeat_region   37626. .37715
                     /rpt_family="(TC)n"
     repeat_region   39025. .39079
                     /rpt_family="(CA)n"
     repeat_region   42932. .42965
                     /rpt_family="(A)n"
     repeat_region   43371. .43392
                     /rpt_family="AT_rich"
     repeat_region   47172. .47236
                     /rpt_family="(GA)n"
     repeat_region   53364. .53538
                     /rpt_family="G-rich"
     repeat_region   54119. .54167
                     /rpt_family="GC_rich"
     repeat_region   54515. .54583
                     /rpt_family="CT-rich"
     repeat_region   55752. .55792
                     /rpt_family="(TG)n"
     repeat_region   56475. .56543
```

```
                     /rpt_family="(CA)n"
     repeat_region   57106. .57145
                     /rpt_family="(CA)n"
     repeat_region   58475. .58548
                     /rpt_family="AluS"
     repeat_region   58595. .58636
                     /rpt_family="(TTTTG)n"
     repeat_region   60111. .60152
                     /rpt_family="G-rich"
     repeat_region   62336. .62360
                     /rpt_family="(CAA)n"
     unsure          64373
                     /note="Probably C."
     repeat_region   65097. .65128
                     /rpt_family="(TTCTCC)n"
     repeat_region   67379. .67404
                     /rpt_family="AT_rich"
     repeat_region   68288. .68314
                     /rpt_family="(T)n"
     repeat_region   69226. .69282
                     /rpt_family="G-rich"
     repeat_region   71515. .71616
                     /rpt_family="(CA)n"
     repeat_region   71778. .71825
                     /rpt_family="(TG)n"
     repeat_region   72721. .72824
                     /rpt_family="C-rich"
     repeat_region   78618. .78664
                     /rpt_family="(TC)n"
     repeat_region   79428. .79511
                     /rpt_family="(GA)n"
     repeat_region   79599. .79672
                     /rpt_family="(TC)n"
     repeat_region   82718. .82748
                     /rpt_family="(T)n"
     repeat_region   83175. .83210

  Query Match             81.6%;  Score 199.2;  DB 10;  Length 112570;
  Best Local Similarity   88.5%;  Pred. No. 2.6e-47;
  Matches  216;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              ||||||||||||||||||||||||| |||||||||||||||| || |  ||||| |||||
Db      86547 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACCGCCGCCGCCGGTAAACT 86606

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              || ||||| |||||||||||||||||||  |||||||| || ||| |  | | ||||||
Db      86607 TTCGCAGGCGGCTCAGCCATTGGCTGACAGAGTCACGTGTCCTTCCCCTCGAGCCCTCCG 86666

Qy        121 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||  |||||||| ||||||||||||||||||||||||||||||||||||| || || ||
Db      86667 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCCTAT 86726

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              |||||||||||||||||||||| ||||| | ||||||||||||||  ||||||||||||
Db      86727 GGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGGCTTCGCAGGACCCAGTCACTCAGTG 86786

Qy        241 ACAG 244
              ||||
Db      86787 ACAG 86790


RESULT 6
AC097087
LOCUS       AC097087               234603 bp    DNA     linear   HTG 30-OCT-2003
DEFINITION  Rattus norvegicus clone CH230-140P4, *** SEQUENCING IN PROGRESS
            ***, 13 unordered pieces.
ACCESSION   AC097087
VERSION     AC097087.12  GI:38044133
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 234603)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 234603)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 234603)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 30, 2003 this sequence version replaced gi:37999244.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
```

```
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GIAV
    Center clone name: CH230-140P4
-------------- Summary Statistics
    Assembly program: Atlas 3.0;
    Consensus quality: 226791 bases at least Q40
    Consensus quality: 228171 bases at least Q30
    Consensus quality: 229344 bases at least Q20
    Estimated insert size: 240235; sum-of-contigs estimation
    Estimated insert size: 274125; agarose-fp estimation
    Quality coverage: 5x in Q20 bases; agarose-fp estimation
    Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1005: contig of 1005 bp in length
*     1006     3091: gap of 2086 bp
*     3092     4962: contig of 1871 bp in length
*     4963     6209: gap of 1247 bp
*     6210     7864: contig of 1655 bp in length
*     7865     8063: gap of 199 bp
*     8064    18114: contig of 10051 bp in length
*    18115    18226: gap of 112 bp
*    18227   219701: contig of 201475 bp in length
*   219702   219801: gap of unknown length
*   219802   221361: contig of 1560 bp in length
*   221362   221551: gap of 190 bp
*   221552   225319: contig of 3768 bp in length
*   225320   225419: gap of unknown length
*   225420   226560: contig of 1141 bp in length
*   226561   226660: gap of unknown length
*   226661   228063: contig of 1403 bp in length
*   228064   228163: gap of unknown length
*   228164   229384: contig of 1221 bp in length
*   229385   229484: gap of unknown length
*   229485   230873: contig of 1389 bp in length
*   230874   230973: gap of unknown length
*   230974   232227: contig of 1254 bp in length
*   232228   232327: gap of unknown length
*   232328   234603: contig of 2276 bp in length.
FEATURES             Location/Qualifiers
     source          1. .234603
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-140P4"
     misc_feature    1. .1005
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3092. .4962
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    6210. .7864
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(8503. .9238)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BZ094084"
     misc_feature    219802. .221361
                     /note="wgs_contig"
ORIGIN

  Query Match             81.6%;  Score 199.2;  DB 2;  Length 234603;
  Best Local Similarity   88.5%;  Pred. No. 2.6e-47;
  Matches  216;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||||||||||||||||||||||| |||||||||||||||| || |  ||||| |||||
Db      25386 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACCGCCGCCGCCGGTAAACT 25445

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              || ||||| |||||||||||||||||||  |||||||| || ||| |  | | ||||||
Db      25446 TTCGCAGGCGGCTCAGCCATTGGCTGACAGAGTCACGTGTCCTTCCCCTCGAGCCCTCCG 25505

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||  |||||||| ||||||||||||||||||||||||||||||||||| || || ||
Db      25506 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCCTAT 25565

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              |||||||||||||||||||||| ||||| | ||||||||||||||  ||||||||||
Db      25566 GGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGGCTTCGCAGGACCCAGTCACTCAGTG 25625

Qy        241 ACAG 244
              ||||
Db      25626 ACAG 25629


RESULT 7
AC122669
LOCUS       AC122669              298853 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-172C22, WORKING DRAFT SEQUENCE, 13
            unordered pieces.
ACCESSION   AC122669
VERSION     AC122669.4  GI:30580875
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 298853)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
```

Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 298853)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 298853)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23265825.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GKZV
Center clone name: CH230-172C22
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 277105 bases at least Q40
Consensus quality: 280204 bases at least Q30
Consensus quality: 282390 bases at least Q20
Estimated insert size: 289519; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
---

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	3941: contig of 3941 bp in length
*	3942	4041: gap of unknown length
*	4042	18898: contig of 14857 bp in length
*	18899	18998: gap of unknown length
*	18999	29431: contig of 10433 bp in length
*	29432	29531: gap of unknown length
*	29532	260502: contig of 230971 bp in length
*	260503	260602: gap of unknown length
*	260603	277410: contig of 16808 bp in length
*	277411	277510: gap of unknown length
*	277511	281394: contig of 3884 bp in length
*	281395	281494: gap of unknown length
*	281495	282773: contig of 1279 bp in length
*	282774	282873: gap of unknown length
*	282874	285229: contig of 2356 bp in length
*	285230	285329: gap of unknown length
*	285330	286945: contig of 1616 bp in length
*	286946	287045: gap of unknown length
*	287046	289575: contig of 2530 bp in length
*	289576	289675: gap of unknown length
*	289676	291071: contig of 1396 bp in length
*	291072	291171: gap of unknown length
*	291172	293319: contig of 2148 bp in length
*	293320	293419: gap of unknown length
*	293420	298853: contig of 5434 bp in length.

FEATURES Location/Qualifiers

source	1. .298853 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-172C22"
misc_feature	1. .1461 /note="wgs_end_extension clone_end:T7"
misc_feature	1123. .119559 /note="clone_boundary clone_end:T7 site:EcoRI end_sequence:BH290517"
misc_feature	18999. .20241 /note="wgs_contig"
misc_feature	complement(144446. .229508) /note="clone_boundary clone_end:Sp6 site:EcoRI end_sequence:BH290519"
misc_feature	268843. .270345 /note="wgs_end_extension clone_end:Sp6"
misc_feature	277511. .278559 /note="wgs_end_extension clone_end:Sp6"

ORIGIN

Query Match 81.6%; Score 199.2; DB 2; Length 298853;
Best Local Similarity 88.5%; Pred. No. 2.5e-47;
Matches 216; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              ||||||||||||||||||||||| |||||||||||||||| || |  ||||| |||||
Db     230392 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACCGCCGCCGCCGGTAAACT 230451

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              || ||||| |||||||||||||||||||  |||||||| || ||| |  | | ||||||
Db     230452 TTCGCAGGCGGCTCAGCCATTGGCTGACAGAGTCACGTGTCCTTCCCCTCGAGCCCTCCG 230511

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
```

```
Db     230512 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCCTAT 230571

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240

Db     230572 GGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGGCTTCGCAGGACCCAGTCACTCAGTG 230631

Qy        241 ACAG 244

Db     230632 ACAG 230635


RESULT 8
AC015583
LOCUS       AC015583               200633 bp    DNA     linear   HTG 14-JAN-2002
DEFINITION  Mus musculus chromosome 6 clone RP23-329P9 map 6, *** SEQUENCING IN
            PROGRESS ***, 5 unordered pieces.
ACCESSION   AC015583
VERSION     AC015583.27  GI:14626313
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200633)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 6, clone RP23-329P9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200633)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 7, 2001 this sequence version replaced gi:12957744.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1082
                Center clone name: 329_P_9
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11709: contig of 11709 bp in length
            *    11710    11809: gap of 100 bp
            *    11810    51644: contig of 39835 bp in length
            *    51645    51744: gap of 100 bp
            *    51745   166213: contig of 114469 bp in length
            *   166214   166313: gap of 100 bp
            *   166314   179517: contig of 13204 bp in length
            *   179518   179617: gap of 100 bp
            *   179618   200633: contig of 21016 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200633
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /map="6"
                     /clone="RP23-329P9"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

  Query Match             75.4%;  Score 184;  DB 2;  Length 200633;
  Best Local Similarity   87.3%;  Pred. No. 7.1e-43;
  Matches  213;  Conservative    0;  Mismatches   30;  Indels    1;  Gaps    1;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60

Db     153775 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACGGCCGCAGCCGGTAAACT 153834

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120

Db     153835 TTCGCAGGCGGCTCAGCCATTGGCTGGTAGAGTCACGTG-TCCTCCCCTCGAGCCCTCCG 153893

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180

Db     153894 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCCTAT 153953

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240

Db     153954 GGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGGCTTCGCAGGATCCAATCACTCAGTG 154013

Qy        241 ACAG 244

Db     154014 ACAG 154017


RESULT 9
AC091106
LOCUS       AC091106               236214 bp    DNA     linear   HTG 17-DEC-2003
DEFINITION  Mus musculus chromosome 6 clone RP23-103L13 map 6, *** SEQUENCING
            IN PROGRESS ***, 9 unordered pieces.
ACCESSION   AC091106
VERSION     AC091106.7  GI:39979566
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 236214)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 6, clone RP23-103L13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236214)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
```

```
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 236214)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 17, 2003 this sequence version replaced gi:31442525.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12520
                Center clone name: 103_L_13
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13453: contig of 13453 bp in length
            *    13454    13553: gap of 100 bp
            *    13554    24211: contig of 10658 bp in length
            *    24212    24311: gap of 100 bp
            *    24312    29735: contig of 5424 bp in length
            *    29736    29835: gap of 100 bp
            *    29836    44483: contig of 14648 bp in length
            *    44484    44583: gap of 100 bp
            *    44584    65367: contig of 20784 bp in length
            *    65368    65467: gap of 100 bp
            *    65468    75780: contig of 10313 bp in length
            *    75781    75880: gap of 100 bp
            *    75881   143236: contig of 67356 bp in length
            *   143237   143336: gap of 100 bp
            *   143337   177855: contig of 34519 bp in length
            *   177856   177955: gap of 100 bp
            *   177956   236214: contig of 58259 bp in length.
FEATURES            Location/Qualifiers
     source         1. .236214
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="6"
                    /map="6"
                    /clone="RP23-103L13"
                    /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

  Query Match           75.4%;  Score 184;  DB 2;  Length 236214;
  Best Local Similarity   87.3%;  Pred. No. 7e-43;
  Matches  213;  Conservative    0;  Mismatches   30;  Indels    1;  Gaps    1;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||||||||||||||||||||||||| |||||||||||||||| || | |||||| |||||
Db      31927 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACGGCCGCAGCCGGTAAACT 31986

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              || ||||| |||||||||||||||  | ||||||||  ||||| |  | | ||||||
Db      31987 TTCGCAGGCGGCTCAGCCATTGGCTGGTAGAGTCACGTG-TCCTCCCCTCGAGCCCTCCG 32045

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||  |||||||| ||||||||||||||||||||||||||||||||||||||||| || || ||
Db      32046 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCCTAT 32105

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              |||||||||||||||||||||| ||||| | |||||||||||||  |   ||||||||||
Db      32106 GGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGGCTTCGCAGGATCCAATCACTCAGTG 32165

Qy        241 ACAG 244
              ||||
Db      32166 ACAG 32169


RESULT 10
AC123672
LOCUS       AC123672              236829 bp    DNA     linear   HTG 16-DEC-2003
DEFINITION  Mus musculus chromosome 6 clone RP23-253E11 map 6, *** SEQUENCING
            IN PROGRESS ***, 9 unordered pieces.
ACCESSION   AC123672
VERSION     AC123672.4  GI:39930763
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 236829)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 6, clone RP23-253E11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236829)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 236829)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 16, 2003 this sequence version replaced gi:28975939.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26388
                Center clone name: 253_E_11
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13701: contig of 13701 bp in length
            *    13702    13801: gap of 100 bp
            *    13802    24723: contig of 10922 bp in length
            *    24724    24823: gap of 100 bp
            *    24824    45150: contig of 20327 bp in length
            *    45151    45250: gap of 100 bp
            *    45251    66059: contig of 20809 bp in length
            *    66060    66159: gap of 100 bp
            *    66160    76371: contig of 10212 bp in length
            *    76372    76471: gap of 100 bp
            *    76472   144414: contig of 67943 bp in length
            *   144415   144514: gap of 100 bp
            *   144515   169981: contig of 25467 bp in length
            *   169982   170081: gap of 100 bp
            *   170082   178409: contig of 8328 bp in length
            *   178410   178509: gap of 100 bp
            *   178510   236829: contig of 58320 bp in length.
FEATURES             Location/Qualifiers
     source          1. .236829
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /map="6"
                     /clone="RP23-253E11"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN
```

```
  Query Match             75.4%;  Score 184;  DB 2;  Length 236829;
  Best Local Similarity   87.3%;  Pred. No. 7e-43;
  Matches  213;  Conservative    0;  Mismatches   30;  Indels    1;  Gaps    1;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||||||||||||||||||||||| |||||||||||||||| || | |||||| ||||||
Db      32786 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACGGCCGCAGCCGGTAAACT 32845

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              || ||||| |||||||||||||||||   | ||||||||  ||||| |  | | ||||||
Db      32846 TTCGCAGGCGGCTCAGCCATTGGCTGGTAGAGTCACGTG-TCCTCCCCTCGAGCCCTCCG 32904

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||  |||||||| ||||||||||||||||||||||||||||||||||| || || ||
Db      32905 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCCTAT 32964

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              |||||||||||||||||||||| ||||| | |||||||||||||  |   ||||||||||
Db      32965 GGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGGCTTCGCAGGATCCAATCACTCAGTG 33024

Qy        241 ACAG 244
              ||||
Db      33025 ACAG 33028


RESULT 11
AX344703/c
LOCUS       AX344703               6118 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 128 from Patent WO0200927.
ACCESSION   AX344703
VERSION     AX344703.1  GI:18492589
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with development genes
  JOURNAL   Patent: WO 0200927-A 128 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6118
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             70.7%;  Score 172.6;  DB 6;  Length 6118;
  Best Local Similarity   81.9%;  Pred. No. 1.7e-39;
  Matches  199;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              | |||||   |||||||| |||||||||| | |||||||||| || || ||| ||||||
Db       1362 AACGCCGAAAATTTAAATACCACTAAAACGATAATCCATCACTACGAAAACCGACAAACT 1303

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              ||| ||  |  |||| |||||  || |||||||||||| ||||||||||| ||||||||||
Db       1302 TTTACAAAAAACTCAACCATTAACTAACACCGTCACGTACCCCTCCTCCAACGTCCTCCG 1243

Qy        121 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              |||||||||||||||||| |||||| |||||||||||||||||||||||||| ||||||
Db       1242 CCCTCCCGCCCCCCCTCTTACGCACTATACATTCATATCATTTTTCTTCTCCGACCCCAT 1183

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              |  || | | ||| ||  |||| |||||||| |||||||  ||||  ||||||| |
Db       1182 AAAAAAAATAAAAAAATTAACACAATCACGCCGAACTTCGCAAAACCAAATCACTCAATA 1123

Qy        241 ACA 243
              |||
Db       1122 ACA 1120
```

```
RESULT 12
AX344702
LOCUS       AX344702               6118 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 127 from Patent WO0200927.
ACCESSION   AX344702
VERSION     AX344702.1  GI:18492588
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with development genes
  JOURNAL   Patent: WO 0200927-A 127 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6118
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match           55.4%;  Score 135.2;  DB 6;  Length 6118;
  Best Local Similarity  72.1%;  Pred. No. 1.4e-28;
  Matches  176;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||| |||||||||||||  | ||||||||||||  || | ||||||| ||| ||| |
Db       4757 AGCGTCGGGGATTTAAATGTTATTAAAACGGTGATTTATTATTGCGGAAGTCGGTAAATT 4816

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              |||| |||||| | ||  ||||| ||| | ||| |||||     |  |  |||||  | ||
Db       4817 TTTGTAGGAGGTTTAGTTATTGGTTGATATCGTTACGTGTTTTTTTTTTAGCGTTTTTCG 4876

Qy        121 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
                |  ||        | ||||| | |||| ||| |||| |||||| || | |||    ||
Db       4877 TTTTTTCGTTTTTTTTTTTTGCGTATTGTATATTTATATTATTTTTTTTTTTCGGTTTTAT 4936

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              |||||||||||||||||||| | ||| ||| |||| |||| ||||  |||| | | ||||
Db       4937 GGAGGAAGTGAGAAAGTTGGTATAGTTACGTCGGGTTTCGTAGGATTAGGTTATTTAGTG 4996

Qy        241 ACAG 244
                | ||
Db       4997 ATAG 5000


RESULT 13
BC032547
LOCUS       BC032547               2529 bp    mRNA    linear   PRI 12-NOV-2003
DEFINITION  Homo sapiens homeo box A1, transcript variant 1, mRNA (cDNA clone
            MGC:45232 IMAGE:5537563), complete cds.
ACCESSION   BC032547
VERSION     BC032547.2  GI:34191750
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2529)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2529)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Aug 25, 2003 this sequence version replaced gi:21595838.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 69 Row: k Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 24497507.
FEATURES             Location/Qualifiers
     source          1. .2529
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:45232 IMAGE:5537563"
                     /tissue_type="Skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_72"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .2529
                     /gene="HOXA1"
                     /note="synonyms: HOX1F, MGC45232"
                     /db_xref="LocusID:3198"
                     /db_xref="MIM:142955"
     CDS             63. .1070
                     /codon_start=1
                     /product="homeobox A1 protein, isoform a"
                     /protein_id="AAH32547.1"
                     /db_xref="GI:21595839"
                     /db_xref="LocusID:3198"
                     /translation="MDNARMNSFLEYPILSSGDSGTCSARAYPSDHRITTPQSCAVSA
                     NSCGGDDRFLVGRGVQIGSPHHHHHHHHHHPQPATYQTSGNLGVSYSHSSCGPSYGSQ
                     NFSAPYSPYALNQEADVSGGYPQCAPAVYSGNLSSPMVQHHHHHQGYAGGAVGSPQYI
                     HHSYGQEHQSLALATYNNSLSPLHASHQEACRSPASETSSPAQTFDWMKVKRNPPKTG
                     KVGEYGYLGQPNAVRTNFTTKQLTELEKEFHFNKYLTRARRVEIAASLQLNETQVKIW
                     FQNRRMKQKKREKEGLLPISPATPPGNDEKAEESSEKSSSSPCVPSPGSSTSDTLTTS
                     H"
```

```
     misc_feature    759. .920
                     /note="homeobox; Region: Homeobox domain"
                     /db_xref="CDD:pfam00046"
ORIGIN

  Query Match             25.4%;  Score 62;  DB 9;  Length 2529;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-07;
  Matches   62;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        183 AGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGAC 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGAC 60

Qy        243 AG 244
              ||
Db         61 AG 62


RESULT 14
MUSERA1A
LOCUS       MUSERA1A                2216 bp    mRNA    linear   ROD 12-JUN-1993
DEFINITION  Mouse ERA-1-993 mRNA, complete cds, and alternate ERA-1-339 mRNA,
            complete cds.
ACCESSION   M22115
VERSION     M22115.1  GI:193047
KEYWORDS    ERA-1 protein; early retinoic acid protein.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2216)
  AUTHORS   LaRosa,G.J. and Gudas,L.J.
  TITLE     Early retinoic acid-induced F9 teratocarcinoma stem cell gene
            ERA-1: alternate splicing creates transcripts for a
            homeobox-containing protein and one lacking the homeobox
  JOURNAL   Mol. Cell. Biol. 8 (9), 3906-3917 (1988)
  MEDLINE   89127233
   PUBMED   2906112
COMMENT     Original source text: Mouse (F9 teratocarcinoma stem cell line)
            mRNA, clone ERA-1-993.
FEATURES             Location/Qualifiers
     source          1. .2216
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS             81. .1076
                     /note="ERA-1-993 protein"
                     /codon_start=1
                     /protein_id="AAA37559.1"
                     /db_xref="GI:309218"
                     /translation="MNSFLEYPILGSGDSGTCSARAYPSDHGITTFQSCAVSANSCGG
                     DDRFLVGRGVQISSPHHHHHHHHHHPQTATYQTSGNLGISYSHSSCGPSYGAQNFSA
                     PYGPYGLNQEADVSGGYPPCAPAVYSGNLSTPMVQHHHHQGYAGGTVGSPQYIHHSY
                     GQEQQTLALATYNNSLSPLHASHQEACRSPASETSSPAQTFDWMKVKRNPPKTGKVGE
                     YGYVGQPNAVRTNFTTKQLTELEKEFHFNKYLTRARRVEIAASLQLNETQVKIWFQNR
                     RMKQKKREKEGLLPISPATPPGSDEKTEESSEKSSPSPSAPSPASSTSDTLTTSH"
     CDS             join(81. .422,626. .685)
                     /note="ERA-1-399 protein"
                     /codon_start=1
                     /protein_id="AAA37558.1"
                     /db_xref="GI:387146"
                     /translation="MNSFLEYPILGSGDSGTCSARAYPSDHGITTFQSCAVSANSCGG
                     DDRFLVGRGVQISSPHHHHHHHHHHPQTATYQTSGNLGISYSHSSCGPSYGAQNFSA
                     PYGPYGLNQEADPPRSLSFPCFRDVFSSADL"
     exon            <81. .422
                     /note="ERA-1-399 protein"
     exon            626. .>685
                     /note="ERA-1-399 protein"
     misc_feature    753. .932
                     /note="ERA-1-993 protein homeobox"
ORIGIN

  Query Match             22.0%;  Score 53.8;  DB 10;  Length 2216;
  Best Local Similarity   89.2%;  Pred. No. 8.6e-05;
  Matches   58;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        180 TGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGT 239
              ||||||||||||||||||||||| ||||| | |||||||||||||| |   |||||||||
Db          1 TGGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGGCTTCGCAGGATCCAATCACTCAGT 60

Qy        240 GACAG 244
              |||||
Db         61 GACAG 65


RESULT 15
MUSHOX161
LOCUS       MUSHOX161               443 bp    DNA     linear   ROD 01-NOV-2000
DEFINITION  Mus musculus Hox1.6 protein (Hox1.6) gene, exon 2.
ACCESSION   M20214 M15928
VERSION     M20214.1  GI:193918
KEYWORDS    homeo box.
SEGMENT     1 of 3
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Baron,A., Featherstone,M.S., Hill,R.E., Hall,A., Galliot,B. and
            Duboule,D.
  TITLE     Hox-1.6: a mouse homeo-box-containing gene member of the Hox-1
            complex
  JOURNAL   EMBO J. 6 (10), 2977-2986 (1987)
  MEDLINE   88082669
   PUBMED   2891503
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone_lib="library of H.Lehrach"
                     /dev_stage="embryo"
     exon            20. .435
                     /gene="Hox1.6"
                     /number=2
ORIGIN

  Query Match             20.9%;  Score 51;  DB 10;  Length 443;
  Best Local Similarity   86.1%;  Pred. No. 0.0006;
  Matches   68;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy        166 CTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGA 225
              |||||| || || |||||||||||||||||||||||||| ||||| ||  ||||||||||
Db          1 CTTCTCTGGTCCTATGGAGGAAGTGAGAAAGTTGGCACGGTCAC-CCATGCTTCGCAGGA 59

Qy        226 CCAGGTCACTCAGTGACAG 244
              |   ||||||||||||||
Db         60 TCCAATCACTCAGTGACAG 78


Search completed: September 18, 2004, 12:08:43
Job time : 1172.45 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 09:40:06 ; Search time 262.692 Seconds
(without alignments)
3945.922 Million cell updates/sec

Title: US-10-614-282-2
Perfect score: 244
Sequence: 1 agcgccggggatttaaatgc.....accaggtcactcagtgacag 244

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	172.6	70.7	6118	6	ABN80111	Abn80111 Human che
	2	135.2	55.4	6118	6	ABN80110	Abn80110 Human che
c	3	38	15.6	302250	6	ABL67703	Abl67703 Oesophagu
	4	37.8	15.5	58857	3	AAA58471	Aaa58471 Nucleotid
	5	37.2	15.2	3050	9	ADB62108	Adb62108 Human cDN
	6	37	15.2	38110	7	AAL53730	Aal53730 Ribonucle
	7	37	15.2	38110	7	AAD49662	Aad49662 Mycobacte
	8	37	15.2	63155	9	ADC85996	Adc85996 Human GPC
	9	37	15.2	110000	4	AAI99682_25	Continuation (26 o
	10	37	15.2	110000	4	AAI99683_25	Continuation (26 o
	11	36	14.8	374	6	ABL81906	Abl81906 Human ova
	12	35.6	14.6	1141	9	ADC87092	Adc87092 Human GPC
	13	35.6	14.6	2179	7	ADA53147	Ada53147 Human cod
	14	35.6	14.6	2818	9	ADB62471	Adb62471 Human cDN
	15	35.6	14.6	3089	9	ADB62760	Adb62760 Human cDN
	16	35.6	14.6	3743	9	ADB62119	Adb62119 Human cDN
	17	35.6	14.6	11706	5	ABA20479	Aba20479 Human ner
	18	35.6	14.6	34875	9	ADD14677	Add14677 Human src
c	19	35.4	14.5	541	9	ADB49520	Adb49520 Primary r
c	20	35	14.3	2363	4	ABL17819	Abl17819 Drosophil
c	21	35	14.3	2876	4	ABL17828	Abl17828 Drosophil
c	22	35	14.3	4363	4	ABL17818	Abl17818 Drosophil
	23	34.6	14.2	3060	2	AAX39667	Aax39667 Renal can
	24	34.6	14.2	4725	5	AAS76375	Aas76375 DNA encod
	25	34.6	14.2	4739	6	ABL63316	Abl63316 Breast ca
	26	34.6	14.2	4739	7	ACC00031	Acc00031 Human Bcr
c	27	34.6	14.2	4756	5	AAS85030	Aas85030 DNA encod
	28	34.6	14.2	4775	5	AAS76377	Aas76377 DNA encod
	29	34.6	14.2	5795	5	AAS85031	Aas85031 DNA encod
	30	34.4	14.1	122888	6	ABK83569	Abk83569 Human cDN
c	31	34	13.9	175737	6	ABK83571	Abk83571 Human cDN
	32	33.6	13.8	1778	6	AAS18963	Aas18963 Wheat ace
c	33	33.4	13.7	2537	6	ABS51803	Abs51803 Human mdd
c	34	33.4	13.7	2550	9	ADD29657	Add29657 Human tum
c	35	33.4	13.7	16161	6	AAD36071	Aad36071 Human adi
	36	33	13.5	152141	7	ACA64961	Aca64961 Human BCR
	37	32.8	13.4	33137	9	ADC86460	Adc86460 Human GPC
	38	32.8	13.4	48840	8	ACH03435	Ach03435 Genomic D
c	39	32.6	13.4	3337	4	ABL27738	Abl27738 Drosophil
c	40	32.4	13.3	315	4	AAF85371	Aaf85371 Degenerat
c	41	32.4	13.3	315	7	ABX12105	Abx12105 Degenerat
c	42	32.4	13.3	4769	3	AAA50439	Aaa50439 Porcine a
	43	32.4	13.3	5349	3	AAA65527	Aaa65527 Porcine B
	44	32.4	13.3	5577	8	ADB12722	Adb12722 Human pol
c	45	32.4	13.3	12001	2	AAQ76213	Aaq76213 HSV L/ST

ALIGNMENTS

RESULT 1
ABN80111/c
ID ABN80111 standard; DNA; 6118 BP.
XX
AC ABN80111;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 128.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 128; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with

CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 6118 BP; 1532 A; 235 C; 1646 G; 2705 T; 0 U; 0 Other;

Query Match 70.7%; Score 172.6; DB 6; Length 6118;
Best Local Similarity 81.9%; Pred. No. 3.6e-42;
Matches 199; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```
Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              | |||||   ||||||||| |||||||||||| | |||||||||| || || ||| ||||||
Db       1362 AACGCCGAAAATTTAAATACCACTAAAACGATAATCCATCACTACGAAAACCGACAAACT 1303

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              ||| ||  |  |||| |||||  || ||||||||||| |||||||||| |||||||||
Db       1302 TTTACAAAAAACTCAACCATTAACTAACACCGTCACGTACCCCTCCTCCAACGTCCTCCG 1243

Qy        121 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              |||||||||||||||||| |||||| |||||||||||||||||||||||||| ||||||
Db       1242 CCCTCCCGCCCCCCCCTCTTACGCACTATACATTCATATCATTTTTCTTCTCCGACCCCAT 1183

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
                |  || | | ||| ||  |||| ||||||||  |||||||  ||||  ||||||| |
Db       1182 AAAAAAAATAAAAAAATTAACACAATCACGCCGAACTTCGCAAAACCAAATCACTCAATA 1123

Qy        241 ACA 243
              |||
Db       1122 ACA 1120
```

RESULT 2
ABN80110
ID ABN80110 standard; DNA; 6118 BP.
XX
AC ABN80110;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 127.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 127; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 6118 BP; 1436 A; 235 C; 1504 G; 2943 T; 0 U; 0 Other;

Query Match 55.4%; Score 135.2; DB 6; Length 6118;
Best Local Similarity 72.1%; Pred. No. 8.9e-31;
Matches 176; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||| |||||||||||||  | ||||||||||  || | |||||||| ||| ||| |
Db       4757 AGCGTCGGGGATTTAAATGTTATTAAAACGGTGATTTATTATTGCGGAAGTCGGTAAATT 4816

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              |||| |||||| | ||  ||||| ||| | ||| |||||    |  |  |||||  | ||
Db       4817 TTTGTAGGAGGTTTAGTTATTGGTTGATATCGTTACGTGTTTTTTTTTTTAGCGTTTTTCG 4876

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
                |  ||       | ||||| | |||| ||| |||| |||||| || | |||    ||
Db       4877 TTTTTTCGTTTTTTTTTTTGCGTATTGTATATTTATATTATTTTTTTTTTCGGTTTTAT 4936

Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              |||||||||||||||||||| | ||| ||| |||| |||| |||| |||| | | ||||
Db       4937 GGAGGAAGTGAGAAAGTTGGTATAGTTACGTCGGGTTTCGTAGGATTAGGTTATTTAGTG 4996

Qy        241 ACAG 244
              | ||
Db       4997 ATAG 5000
```

RESULT 3
ABL67703/c
ID ABL67703 standard; DNA; 302250 BP.
XX
AC ABL67703;
XX
DT 15-MAY-2002 (first entry)
XX
DE Oesophagus cancer related gene sequence SEQ ID NO:6040.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.

```
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 6040; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 302250 BP; 76116 A; 72066 C; 71554 G; 82514 T; 0 U; 0 Other;

  Query Match             15.6%;  Score 38;  DB 6;  Length 302250;
  Best Local Similarity   53.3%;  Pred. No. 1.3;
  Matches   80;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy         91 CGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTAC 150
              | || |   || |||  |   ||||| ||| |  | | |||| |  |   | | ||||
Db     294645 CCTCCCACCCCTCTCACCATTCGTCCCCCGACTCCTCCCCCCGCTCCGCTCCCTCTGTCA 294586

Qy        151 ATTCATATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACG 210
              || |      || ||||   || || ||| ||    |||||  |  |  ||| ||||
Db     294585 GTTGAGCAGTTTCTTCTATCCCTGCACCAGGGTCCTAGTGAACACCTATGCAGAGTCTGA 294526

Qy        211 CCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
               ||| | ||  |  ||||    |  |||
Db     294525 ATTGGCATGGCCTGGGCAGGGACATGTGTG 294496


RESULT 4
AAA58471
ID   AAA58471 standard; DNA; 58857 BP.
XX
AC   AAA58471;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW   BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW   bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW   thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS   Streptomyces verticillus.
XX
FH   Key             Location/Qualifiers
FT   CDS             223. .564
FT                   /*tag=  a
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 30; encodes AAB07556"
FT   CDS             561. .2309
FT                   /*tag=  b
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 29; encodes AAB07557"
FT   CDS             2767. .3486
FT                   /*tag=  c
FT                   /note= "ORF 28; encodes AAB07558"
FT   CDS             3527. .5593
FT                   /*tag=  d
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 27; encodes AAB07559"
FT   CDS             5806. .12294
FT                   /*tag=  e
FT                   /note= "ORF 26; encodes AAB07560"
FT   CDS             12291. .15491
FT                   /*tag=  f
FT                   /note= "ORF 25; encodes AAB07561"
FT   CDS             15488. .21013
FT                   /*tag=  g
FT                   /note= "ORF 24; encodes AAB07562"
FT   CDS             21010. .24666
FT                   /*tag=  h
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 23; encodes AAB07563"
FT   CDS             24663. .32690
FT                   /*tag=  i
FT                   /note= "ORF 22; encodes AAB07564"
```

```
FT   CDS             32893. .34830
FT                   /*tag= j
FT                   /note= "ORF 21; encodes AAB07565"
FT   CDS             34827. .35804
FT                   /*tag= k
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 20; encodes AAB07566"
FT   CDS             35818. .37302
FT                   /*tag= l
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 19; encodes AAB07567"
FT   CDS             37299. .39215
FT                   /*tag= m
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 18; encodes AAB07568"
FT   CDS             39301. .47181
FT                   /*tag= n
FT                   /note= "ORF 17; encodes AAB07569"
FT   CDS             47178. .49985
FT                   /*tag= o
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 16; encodes AAB07570"
FT   CDS             49982. .51001
FT                   /*tag= p
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 15; encodes AAB07571"
FT   CDS             50998. .52386
FT                   /*tag= q
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 14; encodes AAB07572"
FT   CDS             52383. .52946
FT                   /*tag= r
FT                   /note= "ORF 13; encodes AAB07573"
FT   CDS             53018. .54190
FT                   /*tag= s
FT                   /note= "ORF 12; encodes AAB07574"
FT   CDS             54187. .55824
FT                   /*tag= t
FT                   /note= "ORF 11; encodes AAB07575"
FT   CDS             55821. .56093
FT                   /*tag= u
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 10; encodes AAB07576"
FT   CDS             56090. .57586
FT                   /*tag= v
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 9; encodes AAB07577"
FT   CDS             57583. .58857
FT                   /*tag= w
FT                   /note= "ORF 8; encodes AAB07578"
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US000445.
XX
PR   06-JAN-1999;   99US-0115435P.
PR   05-FEB-1999;   99US-0118848P.
PR   05-JAN-2000; 2000US-00477962.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Shen B,  Du L,  Sanchez C,  Chen M,  Edwards DJ;
XX
DR   WPI; 2000-465974/40.
DR   P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR   AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR   AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR   AAB07576, AAB07577, AAB07578.
XX
PT   New bleomycin gene cluster components useful for peptide and/or
PT   polyketide metabolites, especially bleomycin, production and for
PT   chemically modifying biological molecules.
XX
PS   Claim 8; Page 97-136; 162pp; English.
XX
CC   The present sequence represents the BLM (Bleomycin) gene cluster,
CC   containing open reading frames (ORFs) 8-30. The proteins encoded by the
CC   gene cluster are useful for producing peptides and/or polyketide
CC   metabolites, especially bleomycin or bleomycin analogues. They are also
CC   useful for chemically modifying biological molecules to produce branched
CC   methyl groups, and for coupling amino acids and fatty acids. They may be
CC   reacted with an apo-carrier protein and coenzyme A to produce a holo-
CC   carrier protein. The BLM gene cluster or catalytic domains can be used
CC   individually or collectively to produce thiazolidine, thiazoline,
CC   bithiazoline and bithiazoline-containing microbial metabolites. The BLM
CC   gene cluster may also be used to produce sugars
XX
SQ   Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 U; 0 Other;

  Query Match             15.5%;  Score 37.8;  DB 3;  Length 58857;
  Best Local Similarity   50.8%;  Pred. No. 0.88;
  Matches   90;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         44 GCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCC 103
              ||| | | |||||  ||  |   |  |    ||    ||| | ||||    || |||| |
Db      27520 GCCGCACCCGGCGGACCGTGCTGGACGCGGTCGCGCACGGCCGCCACCCCTTCGGGCCGC 27579

Qy        104 TCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTT 163
              || ||   ||||   | ||| | ||| | |||    |||| | |  | | || | | | |
Db      27580 TCGTCTCCCGTCTCGCCCCCGCGCGCACGCCCGGCCGCGCGCCGCTCCTGCAGAGCCTGT 27639

Qy        164 TTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCG 220
              |  | |||| || | |   ||    |||  |  ||      | |||  | | | |
Db      27640 TCGTGCTCCAGCGCGAGTACGGCGACGAGGCGGACGGGTACCGCGCGCTCGCCCTGG 27696


RESULT 5
ADB62108
ID   ADB62108 standard; cDNA; 3050 BP.
XX
AC   ADB62108;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human cDNA encoding clone BRAWH20063010.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1839. .2579
FT                   /*tag= a
FT                   /product= "Clone BRAWH20063010 protein"
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
```

DR WPI; 2003-450961/43.
DR P-PSDB; ADB64078.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 3050 BP; 491 A; 1023 C; 953 G; 583 T; 0 U; 0 Other;

Query Match 15.2%; Score 37.2; DB 9; Length 3050;
Best Local Similarity 56.6%; Pred. No. 0.52;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
Db 2027 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 2086

Qy 89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGT 148
Db 2087 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCCCCCTCCCTCCCTGCGGACTGT 2146

Qy 149 AC 150
Db 2147 GC 2148

RESULT 6
AAL53730
ID AAL53730 standard; DNA; 38110 BP.
XX
AC AAL53730;
XX
DT 07-FEB-2003 (first entry)
XX
DE Ribonuclease P RNase P DNA SEQ ID No 24.
XX
KW Target RNA; target RNA:support-attached test compound; flow cytometry;
KW mass spectrometry; high-throughput screening; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200283837-A1.
XX
PD 24-OCT-2002.
XX
PF 11-APR-2002; 2002WO-US011758.
XX
PR 11-APR-2001; 2001US-0282966P.
XX
PA (PTCT-) PTC THERAPEUTICS INC.
XX
PI Almstead NG;
XX
DR WPI; 2003-075534/07.
XX
PT Identifying a test compound that binds to a target RNA molecule by
PT separating the detectably labeled target RNA:support-attached test
PT compound complex from uncomplexed target RNA molecules and test compounds
PT by flow cytometry.
XX
PS Example; Page 61-79; 131pp; English.
XX
CC The invention relates to a novel method for identifying a test compound
CC that binds to a target RNA molecule comprising separating the detectably
CC labeled target RNA:support-attached test compound complex from
CC uncomplexed target RNA molecules and test compounds. The separating
CC process is carried out by flow cytometry and determining a structure of
CC the type of test compound of the RNA:support-attached test compound
CC complex by mass spectrometry. The method is useful for high-throughput
CC screening of libraries of compounds to identify pharmaceutical leads.
CC This polynucleotide sequence represents a DNA sequence related to the
CC detecting method of the invention
XX
SQ Sequence 38110 BP; 6702 A; 12862 C; 12361 G; 6185 T; 0 U; 0 Other;

Query Match 15.2%; Score 37; DB 7; Length 38110;
Best Local Similarity 52.2%; Pred. No. 1.3;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 45 CGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCT 104
Db 35959 CGGTGGGCCGCTCACATGTGGCGGCGGCAGCGACGTTGGCCGCCTCGGCCCCGTACCGCG 36018

Qy 105 CCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTT 164
Db 36019 CCACCAGCGACTCGGGCAATCCGGCGCCCGATCCGGGGGCCGGCCCAGGGTTCGCCGGTG 36078

Qy 165 TCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGC 201
Db 36079 CGCCGATCAGCGGCAGGTTGCGAGTGCGGCACTTCGC 36115

RESULT 7
AAD49662
ID AAD49662 standard; DNA; 38110 BP.
XX
AC AAD49662;
XX
DT 24-MAR-2003 (first entry)
XX
DE Mycobacterium tuberculosis ribonuclease P (RNAase) rnpB DNA.
XX
KW Amyloidosis; haemophilia; Alzheimer's disease; atherosclerosis; cancer;
KW gigantism; dwarfism; hypothyroidism; hyperthyroidism; cystic fibrosis;
KW autoimmune disorder; aging; inflammation; diabetes; obesity; anorectic;
KW neurodegenerative disorder; Parkinson's disease; gene therapy; virucide;
KW haemostatic; antibacterial; nootropic; neuroprotective; cytostatic;
KW fungicide; ribonuclease P; RNAase; enzyme; gene; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200283953-A1.
XX
PD 24-OCT-2002.
XX
PF 11-APR-2002; 2002WO-US011757.
XX
PR 11-APR-2001; 2001US-0282965P.
XX
PA (PTCT-) PTC THERAPEUTICS INC.

```
XX
PI   Rando R,  Welch E;
XX
DR   WPI; 2003-075561/07.
XX
PT   Identifying a test compound that binds to a target RNA molecule for
PT   treating or preventing amyloidosis, hemophilia, cancer, gigantism,
PT   diabetes, by contacting a detectably labeled target RNA molecule with a
PT   library of test compounds.
XX
PS   Example; Page 70-88; 152pp; English.
XX
CC   The invention relates to a method for identifying a test compound that
CC   binds to a target RNA molecule, which comprises contacting a detectably
CC   labelled target RNA molecule with a library of test compounds under
CC   conditions that permit direct binding of the labelled target RNA to a
CC   member of the library of test compounds so that a detectably labeled
CC   target RNA:test compound complex is formed. The method is useful for
CC   screening libraries of compounds for those that are selectively bind to a
CC   pre-selected target RNA. The compounds are useful for inhibiting the
CC   formation of a specific bound RNA:host cell factor complexes in vivo.
CC   They are also useful for treating or preventing diseases associated with
CC   overproduction or decreased protein function, such as amyloidosis,
CC   haemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism,
CC   dwarfism, hypothyroidism, hyperthyroidism, autoimmune disorders, aging,
CC   inflammation, cystic fibrosis, diabetes, obesity, neurodegenerative
CC   disorders, Parkinson's disease or infections (bacterial, viral, fungal).
CC   The invention is also used in gene therapy. The present sequence is
CC   Mycobacterium tuberculosis ribonuclease P (RNAase) rnpB DNA. This
CC   sequence is used to illustrate the method of the invention
XX
SQ   Sequence 38110 BP; 6702 A; 12862 C; 12361 G; 6185 T; 0 U; 0 Other;

  Query Match             15.2%;  Score 37;  DB 7;  Length 38110;
  Best Local Similarity   52.2%;  Pred. No. 1.3;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         45 CGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCT 104
              ||| | | ||  || | ||  || |||   | | |||||| | | | | | ||| || |
Db      35959 CGGTGGGCCGCTCACATGTGGCGGCGGCAGCGACGTTGGCCGCCTCGGCCCCGTACCGCG 36018

Qy        105 CCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTT 164
              || ||||| |    ||  ||| || |||  ||  | |  | |  ||    |  |   |
Db      36019 CCACCAGCGACTCGGGCAATCCGGCGCCCGATCCGGGGGCCGGCCCAGGGTTCGCCGGTG 36078

Qy        165 TCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGC 201
                | ||  || | |  |||| |  | || ||
Db      36079 CGCCGATCAGCGGCAGGTTGCGAGTGCGGCACTTCGC 36115


RESULT 8
ADC85996
ID   ADC85996 standard; DNA; 63155 BP.
XX
AC   ADC85996;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:449.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC85997.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 449; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 63155 BP; 18875 A; 12104 C; 12350 G; 18904 T; 0 U; 922 Other;

  Query Match             15.2%;  Score 37;  DB 9;  Length 63155;
  Best Local Similarity   64.7%;  Pred. No. 1.6;
  Matches   55;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy         96 CGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCA 155
              | | | |||||||  | |||| |  |||||  | || |||  | |  | | | | |||
Db      60851 CCTCCTCCTCCTCCCCCCTCCTGCTTCCTCCTCCTCCTCCTTCTTCTTATTCTTCCTTCT 60910

Qy        156 TATCATTTTTCTTCTCCGGCCCCAT 180
              | || || |||||||||  | || |
Db      60911 TCTCTTTCTTCTTCTCCTCCTCCTT 60935


RESULT 9
AAI99682_25
Continuation (26 of 45) of AAI99682 from base 2500001 (Mycobacterium tuberculosis strai
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP     Fragment Name      Begin     End
WP     AAI99682_00            1    110000
WP     AAI99682_01       100001    210000
WP     AAI99682_02       200001    310000
WP     AAI99682_03       300001    410000
WP     AAI99682_04       400001    510000
WP     AAI99682_05       500001    610000
WP     AAI99682_06       600001    710000
WP     AAI99682_07       700001    810000
WP     AAI99682_08       800001    910000
WP     AAI99682_09       900001   1010000
WP     AAI99682_10      1000001   1110000
WP     AAI99682_11      1100001   1210000
WP     AAI99682_12      1200001   1310000
WP     AAI99682_13      1300001   1410000
WP     AAI99682_14      1400001   1510000
WP     AAI99682_15      1500001   1610000
WP     AAI99682_16      1600001   1710000
WP     AAI99682_17      1700001   1810000
WP     AAI99682_18      1800001   1910000
WP     AAI99682_19      1900001   2010000
WP     AAI99682_20      2000001   2110000
WP     AAI99682_21      2100001   2210000
WP     AAI99682_22      2200001   2310000
WP     AAI99682_23      2300001   2410000
WP     AAI99682_24      2400001   2510000
WP     AAI99682_25      2500001   2610000
WP     AAI99682_26      2600001   2710000
WP     AAI99682_27      2700001   2810000
WP     AAI99682_28      2800001   2910000
WP     AAI99682_29      2900001   3010000
```

```
WP      AAI99682_30        3000001   3110000
WP      AAI99682_31        3100001   3210000
WP      AAI99682_32        3200001   3310000
WP      AAI99682_33        3300001   3410000
WP      AAI99682_34        3400001   3510000
WP      AAI99682_35        3500001   3610000
WP      AAI99682_36        3600001   3710000
WP      AAI99682_37        3700001   3810000
WP      AAI99682_38        3800001   3910000
WP      AAI99682_39        3900001   4010000
WP      AAI99682_40        4000001   4110000
WP      AAI99682_41        4100001   4210000
WP      AAI99682_42        4200001   4310000
WP      AAI99682_43        4300001   4410000
WP      AAI99682_44        4400001   4411529

  Query Match             15.2%;  Score 37;  DB 4;  Length 110000;
  Best Local Similarity   52.2%;  Pred. No. 1.9;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         45 CGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCT 104
              |||  | | ||  || | ||  || |||   | | ||||| | | | | | ||| || |
Db      23409 CGGTGGGCCGCTCACATGTGGCGGCGGCAGCGACGTTGGCCGCCTCGGCCCCGTACCGCG 23468

Qy        105 CCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTT 164
              || |||||| |    || ||| || |||  || | | | | ||    | | |
Db      23469 CCACCAGCGACTCGGGCAATCCGGCGCCCGATCCGGGGGCCGGCCCAGGGTTCGCCGGTG 23528

Qy        165 TCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGC 201
                | ||  || | |  |||| |  | || ||
Db      23529 CGCCGATCAGCGGCAGGTTGCGAGTGCGGCACTTCGC 23565


RESULT 10
AAI99683_25
Continuation (26 of 44) of AAI99683 from base 2500001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP      Fragment Name      Begin     End
WP      AAI99683_00            1    110000
WP      AAI99683_01       100001    210000
WP      AAI99683_02       200001    310000
WP      AAI99683_03       300001    410000
WP      AAI99683_04       400001    510000
WP      AAI99683_05       500001    610000
WP      AAI99683_06       600001    710000
WP      AAI99683_07       700001    810000
WP      AAI99683_08       800001    910000
WP      AAI99683_09       900001   1010000
WP      AAI99683_10      1000001   1110000
WP      AAI99683_11      1100001   1210000
WP      AAI99683_12      1200001   1310000
WP      AAI99683_13      1300001   1410000
WP      AAI99683_14      1400001   1510000
WP      AAI99683_15      1500001   1610000
WP      AAI99683_16      1600001   1710000
WP      AAI99683_17      1700001   1810000
WP      AAI99683_18      1800001   1910000
WP      AAI99683_19      1900001   2010000
WP      AAI99683_20      2000001   2110000
WP      AAI99683_21      2100001   2210000
WP      AAI99683_22      2200001   2310000
WP      AAI99683_23      2300001   2410000
WP      AAI99683_24      2400001   2510000
WP      AAI99683_25      2500001   2610000
WP      AAI99683_26      2600001   2710000
WP      AAI99683_27      2700001   2810000
WP      AAI99683_28      2800001   2910000
WP      AAI99683_29      2900001   3010000
WP      AAI99683_30      3000001   3110000
WP      AAI99683_31      3100001   3210000
WP      AAI99683_32      3200001   3310000
WP      AAI99683_33      3300001   3410000
WP      AAI99683_34      3400001   3510000
WP      AAI99683_35      3500001   3610000
WP      AAI99683_36      3600001   3710000
WP      AAI99683_37      3700001   3810000
WP      AAI99683_38      3800001   3910000
WP      AAI99683_39      3900001   4010000
WP      AAI99683_40      4000001   4110000
WP      AAI99683_41      4100001   4210000
WP      AAI99683_42      4200001   4310000
WP      AAI99683_43      4300001   4403765

  Query Match             15.2%;  Score 37;  DB 4;  Length 110000;
  Best Local Similarity   52.2%;  Pred. No. 1.9;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         45 CGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCT 104
              |||  | | ||  || | ||  || |||   | | ||||| | | | | | ||| || |
Db      20637 CGGTGGGCCGCTCACATGTGGCGGCGGCAGCGACGTTGGCCGCCTCGGCCCCGTACCGCG 20696

Qy        105 CCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTT 164
              || |||||| |    || ||| || |||  || | | | | ||    | | |
Db      20697 CCACCAGCGACTCGGGCAATCCGGCGCCCGATCCGGGGGCCGGCCCAGGGTTCGCCGGTG 20756

Qy        165 TCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGC 201
                | ||  || | |  |||| |  | || ||
Db      20757 CGCCGATCAGCGGCAGGTTGCGAGTGCGGCACTTCGC 20793


RESULT 11
ABL81906
ID   ABL81906 standard; cDNA; 374 BP.
XX
AC   ABL81906;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:4884.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US017756.
XX
PR   26-MAY-2000; 2000US-0207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 4884; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
```

```
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 374 BP; 69 A; 101 C; 104 G; 100 T; 0 U; 0 Other;

  Query Match             14.8%;  Score 36;  DB 6;  Length 374;
  Best Local Similarity   51.9%;  Pred. No. 0.61;
  Matches   81;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          10 GATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGA 69
               ||| |   |   |  || | |    |||||  |   |  ||  ||  |  |  ||   |
Db         121 GATGTTTGTTGAATGAATAAGTGACTCCATGCCCTTGAGAGTTGGTCAGATGGTGGGAGG 180

Qy          70 GGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGC 129
               ||||   | | |   |||||||  ||     ||||||  ||| || |   ||  || ||  |
Db         181 GGCTGCACGAATCAGTGACACCCTCTAAGTCCCCTCTTCCTGCCTAGCCCAGCCCCCTAC 240

Qy         130 CCCCCCTCTTGCGCACTGTACATTCATATCATTTTT 165
               ||| || ||   |  ||| |    |  |||  ||||
Db         241 CCCACCCCTCTAGGCCTGAAGGGGCTGATCCATTTT 276


RESULT 12
ADC87092
ID   ADC87092 standard; DNA; 1141 BP.
XX
AC   ADC87092;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1545.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87093.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1545; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1141 BP; 128 A; 440 C; 337 G; 236 T; 0 U; 0 Other;

  Query Match             14.6%;  Score 35.6;  DB 9;  Length 1141;
  Best Local Similarity   55.7%;  Pred. No. 1.2;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
               || |  ||| || |||||||   ||     |  |  | || | |   |||  |  ||  |
Db         389 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 448

Qy          89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
               | ||   || || |||||  |    |     || | ||||  || |||  | ||||  |||||
Db         449 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 508

Qy         149 AC 150
               |
Db         509 GC 510


RESULT 13
ADA53147
ID   ADA53147 standard; cDNA; 2179 BP.
XX
AC   ADA53147;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human coding sequence, SEQ ID 715.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   P-PSDB; ADA54786.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 1; SEQ ID NO 715; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 2179 BP; 341 A; 737 C; 661 G; 440 T; 0 U; 0 Other;

  Query Match             14.6%;  Score 35.6;  DB 7;  Length 2179;
  Best Local Similarity   55.7%;  Pred. No. 1.4;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;
```

```
Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||     | | | || | |   |||  |  ||  |
Db       1163 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 1222

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |   |    || | ||||  || |||  | |||| |||||
Db       1223 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 1282

Qy        149 AC 150
               |
Db       1283 GC 1284
```

RESULT 14
ADB62471
ID ADB62471 standard; cDNA; 2818 BP.
XX
AC ADB62471;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone FEBRA20078800.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   24. .644
FT   /*tag= a
FT   /product= "Clone FEBRA20078800 protein"
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB64441.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 2818 BP; 449 A; 956 C; 873 G; 540 T; 0 U; 0 Other;

```
  Query Match           14.6%;  Score 35.6;  DB 9;  Length 2818;
  Best Local Similarity  55.7%;  Pred. No. 1.6;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||     | | | || | |   |||  |  ||  |
Db       1798 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 1857

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |   |    || | ||||  || |||  | |||| |||||
Db       1858 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 1917

Qy        149 AC 150
               |
Db       1918 GC 1919
```

RESULT 15
ADB62760
ID ADB62760 standard; cDNA; 3089 BP.
XX
AC ADB62760;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone NT2NE20117580.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   271. .912
FT   /*tag= a
FT   /product= "Clone NT2NE20117580 protein"
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB64730.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic

PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 3089 BP; 496 A; 1043 C; 957 G; 593 T; 0 U; 0 Other;

Query Match 14.6%; Score 35.6; DB 9; Length 3089;
Best Local Similarity 55.7%; Pred. No. 1.6;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              ||  |  ||| || |||||||   ||     |  |  | || | |   |||  |  ||  |
Db       2066 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 2125

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |   |    || | ||||  || |||  | |||| |||||
Db       2126 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 2185

Qy        149 AC 150
               |
Db       2186 GC 2187
```

Search completed: September 18, 2004, 11:29:57
Job time : 265.692 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:14:16 ; Search time 61.1263 Seconds
(without alignments)
2215.217 Million cell updates/sec

Title: US-10-614-282-2
Perfect score: 244
Sequence: 1 agcgccggggatttaaatgc.....accaggtcactcagtgacag 244

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	37	15.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	2	37	15.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	3	35.4	14.5	7218	1	US-08-232-463-14	Sequence 14, Appl
	4	33.6	13.8	1778	4	US-08-934-386-4	Sequence 4, Appli
c	5	32.4	13.3	315	4	US-09-712-529-6	Sequence 6, Appli
c	6	32.4	13.3	12001	1	US-08-458-568A-11	Sequence 11, Appl
c	7	32.4	13.3	34094	4	US-09-292-034-1	Sequence 1, Appli
	8	32	13.1	4190	3	US-08-924-345-1	Sequence 1, Appli
	9	31	12.7	4895	3	US-09-053-866-1	Sequence 1, Appli
	10	31	12.7	4895	4	US-09-479-130-1	Sequence 1, Appli
	11	31	12.7	4895	4	US-09-472-130A-1	Sequence 1, Appli
	12	30.8	12.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	13	30.4	12.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	14	30.4	12.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	15	30	12.3	832	4	US-09-621-976-2813	Sequence 2813, Ap
	16	29.8	12.2	12565	3	US-09-345-217-3	Sequence 3, Appli
	17	29.6	12.1	3226	3	US-08-870-126-10	Sequence 10, Appl
	18	29.6	12.1	3226	4	US-09-445-247-10	Sequence 10, Appl
c	19	29.2	12.0	2451	4	US-09-252-991A-12219	Sequence 12219, A
	20	29	11.9	432	4	US-09-252-991A-6795	Sequence 6795, Ap
c	21	29	11.9	556	3	US-09-058-489-25	Sequence 25, Appl
	22	29	11.9	816	4	US-09-252-991A-14049	Sequence 14049, A
c	23	29	11.9	933	4	US-09-252-991A-14130	Sequence 14130, A
	24	29	11.9	1230	4	US-09-252-991A-6843	Sequence 6843, Ap
	25	29	11.9	1278	4	US-09-252-991A-14006	Sequence 14006, A
	26	29	11.9	1488	4	US-09-252-991A-6748	Sequence 6748, Ap
c	27	29	11.9	1773	4	US-09-252-991A-6876	Sequence 6876, Ap
	28	28.8	11.8	448	4	US-09-621-976-1871	Sequence 1871, Ap
	29	28.8	11.8	2861	4	US-09-016-434-1103	Sequence 1103, Ap
	30	28.8	11.8	3510	3	US-09-056-105-16	Sequence 16, Appl
	31	28.8	11.8	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	32	28.8	11.8	33529	3	US-09-144-085-3	Sequence 3, Appli
	33	28.6	11.7	11220	3	US-09-105-537-32	Sequence 32, Appl
	34	28.6	11.7	36778	3	US-09-105-537-5	Sequence 5, Appli
	35	28.6	11.7	38506	3	US-09-320-878-19	Sequence 19, Appl
	36	28.6	11.7	38506	4	US-09-141-908-1	Sequence 1, Appli
	37	28.6	11.7	38506	4	US-09-657-440-19	Sequence 19, Appl
	38	28.6	11.7	53526	3	US-08-658-136-2	Sequence 2, Appli
	39	28.6	11.7	53577	3	US-08-658-136-1	Sequence 1, Appli
c	40	28.4	11.6	753	4	US-09-489-039A-1736	Sequence 1736, Ap
	41	28.4	11.6	1227	4	US-09-620-312D-1027	Sequence 1027, Ap
	42	28.4	11.6	1380	4	US-09-620-312D-1026	Sequence 1026, Ap
	43	28.4	11.6	2301	1	US-08-306-691B-23	Sequence 23, Appl
	44	28.4	11.6	2301	4	US-09-167-206-3	Sequence 3, Appli
	45	28.4	11.6	2301	5	PCT-US93-06251-78	Sequence 78, Appl

ALIGNMENTS

RESULT 1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 15.2%; Score 37; DB 3; Length 4403765;
Best Local Similarity 52.2%; Pred. No. 0.6;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy          45 CGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCT 104
               ||| | | ||  || | ||  || |||   | | ||||| | | | | | ||| || |
Db     2520637 CGGTGGGCCGCTCACATGTGGCGGCGGCAGCGACGTTGGCCGCCTCGGCCCCGTACCGCG 2520696

Qy         105 CCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTT 164
               || |||||| |    ||  ||| || |||  ||  | |  | |  ||    | |   |
Db     2520697 CCACCAGCGACTCGGGCAATCCGGCGCCCGATCCGGGGGCCGGCCCAGGGTTCGCCGGTG 2520756

Qy         165 TCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGC 201
                  | ||  || | |  ||||  |  | ||  ||
Db     2520757 CGCCGATCAGCGGCAGGTTGCGAGTGCGGCACTTCGC 2520793
```

RESULT 2
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.

```
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;    LENGTH: 4411529
;    TYPE: DNA
;    ORGANISM: Mycobacterium tuberculosis
;    OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             15.2%;  Score 37;  DB 3;  Length 4411529;
  Best Local Similarity   52.2%;  Pred. No. 0.6;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          45 CGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCT 104
               |||  | | ||  || | ||  || |||   | | ||||| | | | | | ||| || |
Db     2523409 CGGTGGGCCGCTCACATGTGGCGGCGGCAGCGACGTTGGCCGCCTCGGCCCCGTACCGCG 2523468

Qy         105 CCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTT 164
               || |||||| |    ||  ||| || |||  ||  | |  | |  ||    |  |   |
Db     2523469 CCACCAGCGACTCGGGCAATCCGGCGCCCGATCCGGGGGCCGGCCCAGGGTTCGCCGGTG 2523528

Qy         165 TCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGC 201
                  | ||  || |  |  |||| |  | || ||
Db     2523529 CGCCGATCAGCGGCAGGTTGCGAGTGCGGCACTTCGC 2523565


RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             14.5%;  Score 35.4;  DB 1;  Length 7218;
  Best Local Similarity   6.7%;  Pred. No. 0.12;
  Matches    9;  Conservative   85;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          44 GCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCC 103
               | || |||  || | ::::: :      :::  :: ::  ::  : :: :: : : ::::
Db        1052 GAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1111

Qy         104 TCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTT 163
               ::::::  : :::::: ::::::: :::::::::::: : : :: : : ::: : :: :::
Db        1112 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1171

Qy         164 TTCTTCTCCGGCCCC 178
               :::::::::  ::::
Db        1172 YYYYYYYYYYYYYYY 1186


RESULT 4
US-08-934-386-4
; Sequence 4, Application US/08934386
; Patent No. 6306636
;  GENERAL INFORMATION:
;    APPLICANT:  Haselkorn, Robert
;    APPLICANT:  Gornicki, Piotr
;    TITLE OF INVENTION:  Methods for Detecting Nucleic Acid
;    TITLE OF INVENTION:  Segments Encoding Acetyl-CoA Carboxylase
;    NUMBER OF SEQUENCES:  68
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  US
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/934,386
;      FILING DATE:  19-SEP-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  ARSB:521
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1778 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-934-386-4

  Query Match             13.8%;  Score 33.6;  DB 4;  Length 1778;
  Best Local Similarity   55.0%;  Pred. No. 0.25;
  Matches   66;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          42 CTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCC 101
                  | ||  |||| |||| |     || ||||  || | |   |  | | |||||  | |
```

```
Db        349 CCGCCCAAGCAAGCAAGCAGGGCCATGAGGACCAACTACCCGTCGTCCCCGTCTAGACCA 408

Qy        102 CCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCAT 161
              |  || ||| |  || || ||||||| |||||       |      | ||  ||| ||||
Db        409 CACCCCCCACCTGCCCCCACCCTCCCTCCCCCAACTCCTCCATGAATGCACGCATTTCAT 468


RESULT 5
US-09-712-529-6/c
; Sequence 6, Application US/09712529
; Patent No. 6485938
; GENERAL INFORMATION:
;  APPLICANT:  Sheppard, Paul O.
;  APPLICANT:  Bishop, Paul D.
;  APPLICANT:  Whitmore, Theodore E.
;  APPLICANT:  Thompson, Penny P.
;  TITLE OF INVENTION:  Human Zven Proteins
;  FILE REFERENCE:  99-81
;  CURRENT APPLICATION NUMBER: US/09/712,529
;  CURRENT FILING DATE:  2000-11-14
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;  LENGTH: 315
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: This degenerate sequence encodes the amino acid
;   OTHER INFORMATION: sequence of SEQ ID NO:5.
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(315)
;   OTHER INFORMATION: n = A,T,C or G
US-09-712-529-6

  Query Match             13.3%;  Score 32.4;  DB 4;  Length 315;
  Best Local Similarity   33.6%;  Pred. No. 0.3;
  Matches   44;  Conservative   22;  Mismatches   65;  Indels    0;  Gaps    0;

Qy         24 TAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGG 83
              | : |:  :   || || |: |  |:    | : || |  : |:    |: |:|||  :
Db        182 TGRCAYTCYTCNCCYTCNCKNCCNARNGGNGTRCACATNCKNARNCCNCKNARCCANARN 123

Qy         84 CTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGC 143
              :::|   |:  :| |  || | ||:   |   |  | |:: |:|   | |   |:
Db        122 SWDATNGCRCARCANGTNCCNGCNCCRCAYTGNACRTCNCKYTCRCANGCNCCNGTDATN 63

Qy        144 ACTGTACATTC 154
              || | :|| ||
Db         62 ACNGCRCARTC 52


RESULT 6
US-08-458-568A-11/c
; Sequence 11, Application US/08458568A
; Patent No. 5821339
;  GENERAL INFORMATION:
;     APPLICANT:  Schaffer, Priscilla A.
;     APPLICANT:  Yeh, Lily
;     TITLE OF INVENTION:  Compositions and Methods for Treatment of Herpesvirus
;     TITLE OF INVENTION:  Infections
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
;       STREET:  One Liberty Place, 46th floor
;       CITY:  Philadelphia
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19103
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/458,568A
;       FILING DATE:  02-JUNE-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/065,146
;       FILING DATE:  05-MAY-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Leary Ph.D., Kathryn R.
;       REGISTRATION NUMBER:  36,317
;       REFERENCE/DOCKET NUMBER:  DFCI-0029
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (215) 568-3100
;       TELEFAX:  (215) 568-3439
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  12001 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Herpes simplex virus
;       STRAIN:  Herpes Simplex Virus Type 1
US-08-458-568A-11

  Query Match             13.3%;  Score 32.4;  DB 1;  Length 12001;
  Best Local Similarity   64.9%;  Pred. No. 1.5;
  Matches   48;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         72 CTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCC 131
              ||| |||   |  |  | ||  | |   |||| | |||  | |||||||||| | || ||
Db       7690 CTCCGCCCCCGCGTCCCCCCTCCTCCGCCCCCGCGTCCCCCCTCCTCCGCCCCCGCGTCC 7631

Qy        132 CCCCTCTTGCGCAC 145
              |||||| | ||| |
Db       7630 CCCCTCCTCCGCCC 7617


RESULT 7
US-09-292-034-1/c
; Sequence 1, Application US/09292034
; Patent No. 6492343
; GENERAL INFORMATION:
;  APPLICANT: Reddy, P. Seshidhar
;  APPLICANT: Tikoo, Suresh
;  APPLICANT: Babiuk, Lorne
;  TITLE OF INVENTION: PORCINE ADENOVIRUS TYPE 3 GENOME
;  FILE REFERENCE: 293102002400
;  CURRENT APPLICATION NUMBER: US/09/292,034
;  CURRENT FILING DATE: 1999-04-14
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 34094
;   TYPE: DNA
;   ORGANISM: Porcine Adenovirus Type 3
;   FEATURE:
US-09-292-034-1

  Query Match             13.3%;  Score 32.4;  DB 4;  Length 34094;
  Best Local Similarity   48.9%;  Pred. No. 2.3;
  Matches   87;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy         65 CAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCT 124
              | || ||| |       | ||| |||  |||  |||| | ||| | ||       |
Db       3204 CGGGGGGCCCGCAGCGCGTCTGTCACTTTCATCTGCCGCACCTGCGGCCAGGGGTGATTG 3145

Qy        125 CCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCATGGAG 184
                || | | |  |  ||| ||||  |   ||||  | | | ||||  | ||  |
```

```
Db       3144 TCCCCGCACAGGCAGGCGTACTGCTGAGAGGTATCCGTCTCCATCTCGTCCGCCTCCGTA 3085

Qy        185 GAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGAC 242
                |    || |  |||      ||||        ||| || | ||||||||  |   |||

Db       3084 CACTGCAGCAGCTTGTACATCTCACAGGTCTGATCGAAGCAGCAGGTCACACTCAGAC 3027


RESULT 8
US-08-924-345-1
; Sequence 1, Application US/08924345
; Patent No. 6224878
;  GENERAL INFORMATION:
;    APPLICANT:  LEUNG-TACK Patricia
;    APPLICANT:  LEGASTELOIS Isabelle, Christine, Marie-Andree
;    APPLICANT:  AUDONNET Jean-Christophe, Francis
;    APPLICANT:  RIVIERE Michel, Emile, Albert
;    TITLE OF INVENTION:  Mutants and vaccines of the Infectious
;    TITLE OF INVENTION:  Bovine Rhinotracheitis virus
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LARSON AND TAYLOR
;      STREET:  727 SOUTH TWENTY-THIRD STREET
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/924,345
;      FILING DATE:  04-SEP-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/199,172
;      FILING DATE:  09-AUG-1994
;      APPLICATION NUMBER:  FR 92 07930
;      FILING DATE:  26-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SARRO, THOMAS P.
;      REGISTRATION NUMBER:  19396
;      REFERENCE/DOCKET NUMBER:  XI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 920-7200
;      TELEFAX:  (703) 892-8428
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4190 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Type 1 Bovine Herpesvirus
;      STRAIN:  ST
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  172..1311
;      OTHER INFORMATION:  /function= "envelope glycoprotein"
;      OTHER INFORMATION:  /product= "glycoprotein gI"
;      OTHER INFORMATION:  /standard_name= "BHV-1 gI"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1594..3318
;      OTHER INFORMATION:  /function= "envelop glycoprotein"
;      OTHER INFORMATION:  /product= "glycoprotein gE"
;      OTHER INFORMATION:  /standard_name= "BHV-1 gE"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3388..3861
;      OTHER INFORMATION:  /product= "protein US9"
;      OTHER INFORMATION:  /standard_name= "BHV US9"
US-08-924-345-1

  Query Match             13.1%;  Score 32;  DB 3;  Length 4190;
  Best Local Similarity   52.2%;  Pred. No. 1.3;
  Matches   71;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy         42 CTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCC 101
              | |||| || || |   ||   |  ||||    |||  | |      ||   |  | |
Db       3818 CAGCGGCAGGCGCCGTCCTGGCGTTGGAGGTAGTGCCCCGGCCCTGAGGCGGGGCCCGAC 3877

Qy        102 CCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCAT 161
              ||| ||  |  ||||| ||| |||||||| ||   |    ||| |  |     |
Db       3878 TGTCCCCCTCCCCCCTCCCCCCGTCCGCCCCCCGTCCGCCCGCGAGTAAAGGCTGTCTAA 3937

Qy        162 TTTTCTTCTCCGGCCC 177
              |||| | | |  ||||
Db       3938 TTTTTTCCGCACGCCC 3953


RESULT 9
US-09-053-866-1
; Sequence 1, Application US/09053866
; Patent No. 6111075
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Wenfeng
;    APPLICANT:  Presnell, Scott R.
;    APPLICANT:  Yee, David P.
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  PROTEASE-ACTIVATED RECEPTOR
;    TITLE OF INVENTION:  PAR4 (ZCHEMR2)
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/053,866
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leith, Debra K
;      REGISTRATION NUMBER:  32,619
;      REFERENCE/DOCKET NUMBER:  98-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6674
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4895 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  176...1330
;      OTHER INFORMATION:
US-09-053-866-1
```

```
  Query Match             12.7%;  Score 31;  DB 3;  Length 4895;
  Best Local Similarity   56.3%;  Pred. No. 2.9;
  Matches   58;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         98 TGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATA 157
              | |||   || | | |  | |   ||| ||  || | | | |||   |||  | ||| |
Db       3108 TCCCCTGACTGCTGGGGACCCTCGCCTTCCCTCCTCTCACCTGCAGGCTGATCCTTCTTT 3167

Qy        158 TCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGG 200
              ||| |||   ||   | | ||| |||  | ||| || | | ||
Db       3168 TCACTTTCTGTCAATGTCACCAGGGATAAGGTGGGACAATGGG 3210


RESULT 10
US-09-479-130-1
; Sequence 1, Application US/09479130
; Patent No. 6436400
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Wenfeng
;    APPLICANT:  Presnell, Scott R.
;    APPLICANT:  Yee, David P.
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  PROTEASE-ACTIVATED RECEPTOR
;    TITLE OF INVENTION:  PAR4 (ZCHEMR2)
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/479,130
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leith, Debra K
;      REGISTRATION NUMBER:  32,619
;      REFERENCE/DOCKET NUMBER:  98-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6674
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4895 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  176...1330
;      OTHER INFORMATION:
US-09-479-130-1

  Query Match             12.7%;  Score 31;  DB 4;  Length 4895;
  Best Local Similarity   56.3%;  Pred. No. 2.9;
  Matches   58;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         98 TGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATA 157
              | |||   || | | |  | |   ||| ||  || | | | |||   |||  | ||| |
Db       3108 TCCCCTGACTGCTGGGGACCCTCGCCTTCCCTCCTCTCACCTGCAGGCTGATCCTTCTTT 3167

Qy        158 TCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGG 200
              ||| |||   ||   | | ||| |||  | ||| || | | ||
Db       3168 TCACTTTCTGTCAATGTCACCAGGGATAAGGTGGGACAATGGG 3210


RESULT 11
US-09-472-130A-1
; Sequence 1, Application US/09472130A
; Patent No. 6473765
; GENERAL INFORMATION:
;  APPLICANT: Xu, Wenfeng
;  APPLICANT:  Presnell, Scott R.
;  APPLICANT:  Yee, David P.
;  APPLICANT:  Foster, Donald C.
;  TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR PAR4
;  TITLE OF INVENTION:  (ZCHEMR2)
;  FILE REFERENCE: 98-10D2
;  CURRENT APPLICATION NUMBER: US/09/472,130A
;  CURRENT FILING DATE:  2000-01-07
;  PRIOR APPLICATION NUMBER: US 09/053,866
;  PRIOR FILING DATE: 1998-04-01
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 4895
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (176)...(1330)
US-09-472-130A-1

  Query Match             12.7%;  Score 31;  DB 4;  Length 4895;
  Best Local Similarity   56.3%;  Pred. No. 2.9;
  Matches   58;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         98 TGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATA 157
              | |||   || | | |  | |   ||| ||  || | | | |||   |||  | ||| |
Db       3108 TCCCCTGACTGCTGGGGACCCTCGCCTTCCCTCCTCTCACCTGCAGGCTGATCCTTCTTT 3167

Qy        158 TCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGG 200
              ||| |||   ||   | | ||| |||  | ||| || | | ||
Db       3168 TCACTTTCTGTCAATGTCACCAGGGATAAGGTGGGACAATGGG 3210


RESULT 12
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature

```
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             12.6%;  Score 30.8;  DB 4;  Length 1664976;
  Best Local Similarity   61.0%;  Pred. No. 40;
  Matches   50;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         96 CGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCA 155
              | | || | || |||| |||| | || |||||| |||||    ||||      | | |
Db     919539 CCTCCCTCGCCCCCAGATTCCTTCTCCGTCCCGCACCCCCGTGAGCGCCGACGTCCTGCC 919598

Qy        156 TATCATTTTTCTTCTCCGGCCC 177
              || | ||  ||   ||||| ||
Db     919599 TACCGTTGCTCCCTTCCGGGCC 919620


RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             12.5%;  Score 30.4;  DB 3;  Length 4403765;
  Best Local Similarity   50.7%;  Pred. No. 64;
  Matches   73;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         88 CACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTG 147
              |||||  | |||||  |||||  |      | | ||  | ||||| || |      | |||
Db    2062823 CACCGGAATGTGCCGTTCCTCGACAAGACGCAGCAATGCCGCCTCCGCATGGCCCCACCA 2062764

Qy        148 TACATTCATATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTC 207
              || ||  || |    |  | |||| | |  ||  |  | || |  | |  ||
Db    2062763 GACGTTGGTACCTGCCATGATGACCGGACGCTGGGCCGTCGAAAGCAGGCCCGCCGCCCG 2062704

Qy        208 ACGCCGGGCTTCGCAGGACCAGGT 231
                | |||| |||| ||    |||
Db    2062703 GTCCAGGGCGTCGCCGGCTGGGGT 2062680


RESULT 14
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             12.5%;  Score 30.4;  DB 3;  Length 4411529;
  Best Local Similarity   50.7%;  Pred. No. 64;
  Matches   73;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         88 CACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTG 147
              |||||  | |||||  |||||  |      | | ||  | ||||| || |      | |||
Db    2065516 CACCGGAATGTGCCGTTCCTCGACAAGACGCAGCAATGCCGCCTCCGCATGGCCCCACCA 2065457

Qy        148 TACATTCATATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTC 207
              || ||  || |    |  | |||| | |  ||  |  | || |  | |  ||
Db    2065456 GACGTTGGTACCTGCCATGATGACCGGACGCTGGGCCGTCGAAAGCAGGCCCGCCGCCCG 2065397

Qy        208 ACGCCGGGCTTCGCAGGACCAGGT 231
                | |||| |||| ||    |||
Db    2065396 GTCCAGGGCGTCGCCGGCTGGGGT 2065373


RESULT 15
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             12.3%;  Score 30;  DB 4;  Length 832;
  Best Local Similarity   9.6%;  Pred. No. 2.8;
  Matches   18;  Conservative   95;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         11 ATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAG 70
              ::|:: :| : :  :::: :: :::::::   : | :: : :   :::: ::  :  ::
Db        357 WKTWWYWTTYTWTRMWWMKKKARRWYYWWKSTYACASRYRKYTWGWWWYMWKRMMSTRWY 298

Qy         71 GCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCC 130
               : |::||:: |:   :: : :  : ::      ::: : ::: | ::: : : :  ::|
Db        297 CYMCWKCCMYRGRRCAWYTMARGRMWSYAWGKWKSMRSAMSMCTRMYYKKGSTYWTMKCT 238

Qy        131 CCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTG 190
              |   |: ::::  ::::: |:  ::  ::  :| |  :: :  : :: :  :|::::|
Db        237 CATWCYWYWKYWKRMWSKTCWSGSRGGYMTSYTSTRSYSMYWASWMYTMCWWWGRWWSTY 178

Qy        191 AGAAAGTT 198
              : :|:|::
Db        177 WYMAWGKK 170


Search completed: September 18, 2004, 13:19:20
Job time : 83.1263 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 12:08:47 ; Search time 269.259 Seconds
(without alignments)
4573.952 Million cell updates/sec

Title: US-10-614-282-2
Perfect score: 244
Sequence: 1 agcgccggggatttaaatgc.....accaggtcactcagtgacag 244

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	244	100.0	244	17	US-10-614-282-2	Sequence 2, Appli
c	2	38	15.6	302250	9	US-09-962-832-154	Sequence 154, App
c	3	37.2	15.2	288	13	US-10-424-599-116935	Sequence 116935,
	4	37.2	15.2	3050	16	US-10-104-047-262	Sequence 262, App
	5	37	15.2	63155	16	US-10-292-798-449	Sequence 449, App
c	6	36.2	14.8	89829	17	US-10-322-281-359	Sequence 359, App
	7	36	14.8	374	9	US-09-867-701-4884	Sequence 4884, Ap
	8	35.6	14.6	1141	15	US-10-017-161-1889	Sequence 1889, Ap
	9	35.6	14.6	1141	16	US-10-292-798-1545	Sequence 1545, Ap
	10	35.6	14.6	2179	16	US-10-094-749-715	Sequence 715, App
	11	35.6	14.6	2818	16	US-10-104-047-625	Sequence 625, App
	12	35.6	14.6	3089	16	US-10-104-047-914	Sequence 914, App
	13	35.6	14.6	3743	16	US-10-104-047-273	Sequence 273, App
	14	35.6	14.6	34875	17	US-10-775-169-316	Sequence 316, App
c	15	35.4	14.5	900	13	US-10-424-599-141831	Sequence 141831,
	16	34.6	14.2	1616	17	US-10-437-963-78180	Sequence 78180, A
c	17	34.6	14.2	1616	17	US-10-437-963-78180	Sequence 78180, A
	18	34.6	14.2	4739	9	US-09-954-531-586	Sequence 586, App
	19	34.6	14.2	4739	13	US-10-193-651-22	Sequence 22, Appl
c	20	34.2	14.0	104514	13	US-10-087-192-1981	Sequence 1981, Ap
c	21	34	13.9	536	9	US-09-796-692-5305	Sequence 5305, Ap
c	22	34	13.9	536	15	US-10-040-862-5305	Sequence 5305, Ap
c	23	34	13.9	536	16	US-10-057-475B-5305	Sequence 5305, Ap
c	24	34	13.9	536	16	US-10-154-884B-5305	Sequence 5305, Ap
c	25	34	13.9	536	17	US-10-764-324-5305	Sequence 5305, Ap
	26	33.4	13.7	1681	13	US-10-425-114-27309	Sequence 27309, A
c	27	33.4	13.7	2537	13	US-10-466-164-25	Sequence 25, Appl
c	28	33.4	13.7	16161	10	US-09-957-956-8	Sequence 8, Appli
	29	33.2	13.6	731	13	US-10-027-632-153404	Sequence 153404,
	30	33.2	13.6	731	16	US-10-027-632-153404	Sequence 153404,
	31	32.8	13.4	875	17	US-10-767-701-12676	Sequence 12676, A
c	32	32.8	13.4	1004	17	US-10-437-963-26175	Sequence 26175, A
c	33	32.8	13.4	1800	16	US-10-369-493-44508	Sequence 44508, A
	34	32.8	13.4	1950	17	US-10-437-963-42029	Sequence 42029, A
	35	32.8	13.4	33137	16	US-10-292-798-913	Sequence 913, App
	36	32.8	13.4	48841	10	US-09-844-653-32	Sequence 32, Appl
	37	32.6	13.4	540	13	US-10-424-599-118898	Sequence 118898,
c	38	32.4	13.3	315	15	US-10-212-355-6	Sequence 6, Appli
c	39	32.4	13.3	315	15	US-10-212-201-6	Sequence 6, Appli
c	40	32.4	13.3	315	17	US-10-680-755A-6	Sequence 6, Appli
c	41	32.4	13.3	315	17	US-10-680-800A-6	Sequence 6, Appli
	42	32.4	13.3	492	13	US-10-424-599-59619	Sequence 59619, A
	43	32.4	13.3	1959	16	US-10-260-238-1522	Sequence 1522, Ap
	44	32.4	13.3	2097	16	US-10-369-493-28211	Sequence 28211, A
	45	32.4	13.3	2115	16	US-10-369-493-30970	Sequence 30970, A

ALIGNMENTS

RESULT 1
US-10-614-282-2
; Sequence 2, Application US/10614282
; Publication No. US20040082034A1
; GENERAL INFORMATION:
; APPLICANT: LEE, DUNG-FANG
; APPLICANT: JUANG, JYH-LYH
; TITLE OF INVENTION: INTERNAL RIBOSOME ENTRY SITE OF THE LABIAL GENE FOR
; TITLE OF INVENTION: PROTEIN EXPRESSION
; FILE REFERENCE: 08842.0001-00000
; CURRENT APPLICATION NUMBER: US/10/614,282
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: 60/394,270
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 244
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-614-282-2

Query Match 100.0%; Score 244; DB 17; Length 244;
Best Local Similarity 100.0%; Pred. No. 5.5e-75;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120

Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCCAT 180
```

```
Qy        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTG 240

Qy        241 ACAG 244
              ||||
Db        241 ACAG 244
```

RESULT 2
US-09-962-832-154/c
; Sequence 154, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 154
; LENGTH: 302250
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-832-154

Query Match 15.6%; Score 38; DB 9; Length 302250;
Best Local Similarity 53.3%; Pred. No. 0.067;
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy         91 CGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTAC 150
              | || |   || |||  |   ||||| ||| |  | | |||| |  |   | | ||||
Db     294645 CCTCCCACCCCTCTCACCATTCGTCCCCCGACTCCTCCCCCCGCTCCGCTCCCTCTGTCA 294586

Qy        151 ATTCATATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACG 210
               || |     || ||||   || || ||| ||    |||||  |  |  ||| ||||
Db     294585 GTTGAGCAGTTTCTTCTATCCCTGCACCAGGGTCCTAGTGAACACCTATGCAGAGTCTGA 294526

Qy        211 CCGGGCTTCGCAGGACCAGGTCACTCAGTG 240
               ||| | ||  |  ||||   |  |||
Db     294525 ATTGGCATGGCCTGGGCAGGGACATGTGTG 294496
```

RESULT 3
US-10-424-599-116935/c
; Sequence 116935, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 116935
; LENGTH: 288
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(288)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_76603C.1
US-10-424-599-116935

Query Match 15.2%; Score 37.2; DB 13; Length 288;
Best Local Similarity 54.5%; Pred. No. 0.014;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy         59 CTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTC 118
              ||| |||  | ||||     |     ||  |  | || |    | ||| ||| |  |
Db        224 CTTGTGCGTGTGGCTGGTGCTGCCNCTNTCCNCNTCCCCGTGCTCTCTTCCCGTTTGGGT 165

Qy        119 CGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCCC 178
              |  || ||| |||||  ||| | |  ||      ||| | |  ||||| ||||| |  |||
Db        164 CTTCCCCCCCCCCCCTTTCTNGTGGTCTTCTTCTTCTTCTNNTTTTTTTTCTCTGCTCCC 105

Qy        179 A 179
              |
Db        104 A 104
```

RESULT 4
US-10-104-047-262
; Sequence 262, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 262
; LENGTH: 3050
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-262

Query Match 15.2%; Score 37.2; DB 16; Length 3050;
Best Local Similarity 56.6%; Pred. No. 0.029;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||      |  |  | || | |   ||| |  || |
Db       2027 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 2086

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |   |    || | |||||  || |||  | |||| |||||
Db       2087 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCCCCCTCCCTCCCTGCGGACTGT 2146

Qy        149 AC 150
              ||
Db       2147 GC 2148
```

RESULT 5
US-10-292-798-449
; Sequence 449, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18

```
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 449
;   LENGTH: 63155
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(63155)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(246)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1119)..(1317)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7750)..(8041)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10301)..(10493)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10581)..(10762)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (11390)..(11601)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (15029)..(15153)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (61937)..(62955)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (3192)..(3192)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (16147)..(16246)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (18790)..(18889)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (21048)..(21048)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (21066)..(21165)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (33549)..(33648)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (50114)..(50213)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (54694)..(54793)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57161)..(57164)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57166)..(57168)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57175)..(57175)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57179)..(57180)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57182)..(57186)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57188)..(57191)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57198)..(57198)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57223)..(57322)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (59045)..(59144)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (60650)..(60749)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-449

  Query Match           15.2%;  Score 37;  DB 16;  Length 63155;
  Best Local Similarity  64.7%;  Pred. No. 0.091;
  Matches   55;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy         96 CGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCA 155
              | | | |||||||  | |||| | |||||  | || |||  | |  | | | | |||
Db      60851 CCTCCTCCTCCTCCCCCCTCCTGCTTCCTCCTCCTCCTCCTTCTTCTTATTCTTCCTTCT 60910

Qy        156 TATCATTTTTCTTCTCCGGCCCCAT 180
              | || || |||||||||  | || |
Db      60911 TCTCTTTCTTCTTCTCCTCCTCCTT 60935


RESULT 6
US-10-322-281-359/c
; Sequence 359, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc S. Malandro
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 529452001000
;  CURRENT APPLICATION NUMBER: US/10/322,281
;  CURRENT FILING DATE:  2002-12-17
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 359
;   LENGTH: 89829
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(89829)
;   OTHER INFORMATION: n = A,T,C or G
US-10-322-281-359
```

Query Match 14.8%; Score 36.2; DB 17; Length 89829;
Best Local Similarity 56.2%; Pred. No. 0.19;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy         57 AACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCC 116
              | |||  | ||||||   ||     ||  ||| ||  | |   |||||||  |  | |||
Db      66724 ACCTTGGGGAGGAGGGGGAGGAGGAGGAGGACCCCCCCTCCCCTCCCTCCTGCCCCTTCC 66665

Qy        117 TCCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCC 176
              |||  | ||||  | |||||  |   | | |  |||        ||  | | ||  ||||  ||
Db      66664 TCCTCTCTCCACCTCCCCCGCCCCCTCCCCTTACTCCTCCTTCTCTCTGCTCCTCCCCCC 66605

Qy        177 C 177
              |
Db      66604 C 66604
```

RESULT 7
US-09-867-701-4884
; Sequence 4884, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4884
; LENGTH: 374
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4884

Query Match 14.8%; Score 36; DB 9; Length 374;
Best Local Similarity 51.9%; Pred. No. 0.039;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy          10 GATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGA 69
               |||  |  |  |  || | |   |||||  |   |  ||  ||  |  |  ||   |
Db         121 GATGTTTGTTGAATGAATAAGTGACTCCATGCCCTTGAGAGTTGGTCAGATGGTGGGAGG 180

Qy          70 GGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGC 129
               ||||   | | |   |||||||  ||    ||||||  |||  ||  |   ||  ||  ||  |
Db         181 GGCTGCACGAATCAGTGACACCCTCTAAGTCCCCTCTTCCTGCCTAGCCCAGCCCCCTAC 240

Qy         130 CCCCCCTCTTGCGCACTGTACATTCATATCATTTTT 165
               |||  ||  ||   |  |||  |   |  |||  ||||
Db         241 CCCACCCCTCTAGGCCTGAAGGGGCTGATCCATTTT 276
```

RESULT 8
US-10-017-161-1889
; Sequence 1889, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1889
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1141)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(941)
US-10-017-161-1889

Query Match 14.6%; Score 35.6; DB 15; Length 1141;
Best Local Similarity 55.7%; Pred. No. 0.077;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Qy          29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
               || |  ||| || |||||||   ||     |  |  | || | |   |||  |  ||  |
Db         389 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCCTCCCTCCC 448

Qy          89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
               | ||   || || |||||  |   |    || | ||||  || |||  | |||| |||||
Db         449 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 508

Qy         149 AC 150
               |
Db         509 GC 510
```

RESULT 9
US-10-292-798-1545
; Sequence 1545, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1545
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1141)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(941)
US-10-292-798-1545

Query Match 14.6%; Score 35.6; DB 16; Length 1141;
Best Local Similarity 55.7%; Pred. No. 0.077;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Qy          29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
               || |  ||| || |||||||   ||     |  |  | || | |   |||  |  ||  |
Db         389 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCCTCCCTCCC 448

Qy          89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
               | ||   || || |||||  |   |    || | ||||  || |||  | |||| |||||
Db         449 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 508

Qy         149 AC 150
```

```
Db        509 GC 510


RESULT 10
US-10-094-749-715
; Sequence 715, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 715
;   LENGTH: 2179
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-715

  Query Match            14.6%;  Score 35.6;  DB 16;  Length 2179;
  Best Local Similarity   55.7%;  Pred. No. 0.095;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||      |  |  | || | |   |||  |  ||  |
Db       1163 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 1222

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |    |    || | ||||  || |||  | |||| |||||
Db       1223 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 1282

Qy        149 AC 150
              ||
Db       1283 GC 1284


RESULT 11
US-10-104-047-625
; Sequence 625, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 625
;   LENGTH: 2818
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-625

  Query Match            14.6%;  Score 35.6;  DB 16;  Length 2818;
  Best Local Similarity   55.7%;  Pred. No. 0.1;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||      |  |  | || | |   |||  |  ||  |
Db       1798 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 1857

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |    |    || | ||||  || |||  | |||| |||||
Db       1858 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 1917

Qy        149 AC 150
              ||
Db       1918 GC 1919


RESULT 12
US-10-104-047-914
; Sequence 914, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 914
;   LENGTH: 3089
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-914

  Query Match            14.6%;  Score 35.6;  DB 16;  Length 3089;
  Best Local Similarity   55.7%;  Pred. No. 0.11;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||      |  |  | || | |   |||  |  ||  |
Db       2066 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 2125

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |    |    || | ||||  || |||  | |||| |||||
Db       2126 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 2185

Qy        149 AC 150
              ||
Db       2186 GC 2187


RESULT 13
US-10-104-047-273
; Sequence 273, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
```

```
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 273
;   LENGTH: 3743
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-273

  Query Match             14.6%;  Score 35.6;  DB 16;  Length 3743;
  Best Local Similarity   55.7%;  Pred. No. 0.11;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||      |  |  | ||| | |   |||  |  ||  |
Db       2720 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 2779

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |   |    || | ||||  || |||  | |||| |||||
Db       2780 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 2839

Qy        149 AC 150
              |
Db       2840 GC 2841


RESULT 14
US-10-775-169-316
; Sequence 316, Application US/10775169
; Publication No. US20040175743A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Burczynski, Michael
;  APPLICANT:  Twine, Natalie
;  APPLICANT:  Dorner, Andrew
;  APPLICANT:  Trepicchio, William
;  TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
;  FILE REFERENCE: AM101080 (031896-013000)
;  CURRENT APPLICATION NUMBER: US/10/775,169
;  CURRENT FILING DATE:  2004-02-11
;  NUMBER OF SEQ ID NOS: 5278
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 316
;   LENGTH: 34875
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-169-316

  Query Match             14.6%;  Score 35.6;  DB 17;  Length 34875;
  Best Local Similarity   55.7%;  Pred. No. 0.23;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         29 CGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATTGGCTGAC 88
              || |  ||| || |||||||   ||      |  |  | ||| | |   |||  |  ||  |
Db      18170 CGCTCCTCCCTCCCTGCGGACCGCGCTCGGCCCTCCCTGGGGACCGCGCCCCTCCCTCCC 18229

Qy         89 ACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCCCCCCCCTCTTGCGCACTGT 148
              | ||   || || |||||  |   |    || | ||||  || |||  | |||| |||||
Db      18230 AGCGGACCGCGCTCCTCCCTCCCTGGGGACCTCGCTCCTCCCTCCCTCCCTGCGGACTGT 18289

Qy        149 AC 150
              |
Db      18290 GC 18291


RESULT 15
US-10-424-599-141831/c
; Sequence 141831, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 141831
;   LENGTH: 900
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_99085C.1
US-10-424-599-141831

  Query Match             14.5%;  Score 35.4;  DB 13;  Length 900;
  Best Local Similarity   57.8%;  Pred. No. 0.084;
  Matches   63;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         22 ACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACTTTTGCAGGAGGCTCAGCCATT 81
              || || |  |  ||||     |||  |||| ||     |  || |  ||| || |  ||
Db        479 ACGAAGAACGCAATCCCAATTTGCATAAGCAGGACCCCAATTTCCTGAGCCTGAATCAGG 420

Qy         82 GGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCGCCCTCCCGCC 130
              ||||| || || |||  || || ||   | |  |  ||||||| || ||
Db        419 GGCTGCCAACGCCACCGGCGCCACCGTGACCACCACCCGCCCTTCCTCC 371


Search completed: September 18, 2004, 15:14:59
Job time : 271.259 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 10:33:16 ; Search time 2061.12 Seconds
(without alignments)
3535.155 Million cell updates/sec

Title: US-10-614-282-2
Perfect score: 244
Sequence: 1 agcgccggggatttaaatgc.....accaggtcactcagtgacag 244

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	125.8	51.6	421	10	BE649209	BE649209 UI-M-BH2.
	2	124.2	50.9	343	10	AW121149	AW121149 UI-M-BH2.
	3	88	36.1	625	12	BM838150	BM838150 K-EST0114
	4	80	32.8	2475	11	AK083575	AK083575 Mus muscu
	5	77	31.6	428	9	AA199907	AA199907 zq52a07.r
	6	76.4	31.3	604	10	BB655265	BB655265 BB655265
	7	73	29.9	432	10	BB846631	BB846631 BB846631
	8	63	25.8	1104	12	BM462913	BM462913 AGENCOURT
	9	62	25.4	851	9	AU127876	AU127876 AU127876
	10	55	22.5	773	9	AU132057	AU132057 AU132057
	11	54	22.1	471	9	AA070261	AA070261 zm69c05.r
	12	50	20.5	640	9	AU128573	AU128573 AU128573
	13	44.6	18.3	661	14	CA415315	CA415315 UI-H-EZ0-
	14	44	18.0	678	9	AL650594	AL650594 AL650594
	15	43	17.6	550	9	AU129261	AU129261 AU129261
	16	43	17.6	683	9	AU128985	AU128985 AU128985
	17	41.4	17.0	913	13	BX408305	BX408305 BX408305
c	18	41.2	16.9	793	12	BG344149	BG344149 HVSMEg000
	19	41	16.8	604	9	AU128022	AU128022 AU128022
c	20	40.6	16.6	1201	29	CNS01668	AL106346 Drosophil
c	21	39.2	16.1	465	14	CD333977	CD333977 StrPu536.
	22	39	16.0	619	12	BI886285	BI886285 ZF637-1-0
c	23	38.2	15.7	899	13	BX349846	BX349846 BX349846
c	24	38	15.6	983	14	CA976262	CA976262 AGENCOURT
c	25	37.8	15.5	686	29	CNS0333T	AL225650 Tetraodon
c	26	37.8	15.5	1201	9	AL535176	AL535176 AL535176
c	27	37.4	15.3	853	29	AG138149	AG138149 Pan trogl
c	28	37.4	15.3	1033	28	AQ893007	AQ893007 HS_4832_A
	29	37.4	15.3	1101	29	CNS016XU	AL107340 Drosophil
c	30	37.2	15.2	684	12	BI957311	BI957311 HVSMEn000
c	31	37	15.2	907	29	CG276221	CG276221 OG1DR10TV
c	32	37	15.2	1004	28	BZ559376	BZ559376 pacs2-164
	33	36.8	15.1	819	29	CC907659	CC907659 t036e05ba
	34	36.8	15.1	1021	28	CC225169	CC225169 CH261-114
	35	36.6	15.0	469	28	BH254220	BH254220 SALK_0161
	36	36.6	15.0	483	10	AW447419	AW447419 88968 MAR
c	37	36.6	15.0	602	29	CNS03K0U	AL247575 Tetraodon
	38	36.6	15.0	603	12	BI074777	BI074777 IP1_15_B1
	39	36.6	15.0	620	14	CB462398	CB462398 722478 MA
c	40	36.6	15.0	621	12	BI064787	BI064787 pgf1n.pk0
c	41	36.6	15.0	1605	10	BF138648	BF138648 601781826
c	42	36.4	14.9	415	13	BY447104	BY447104 BY447104
	43	36.4	14.9	469	12	BM508777	BM508777 ii39g08.y
c	44	36.4	14.9	866	9	AL525171	AL525171 AL525171
	45	36.4	14.9	893	10	BF122415	BF122415 601759708

ALIGNMENTS

RESULT 1
BE649209/c
LOCUS BE649209 421 bp mRNA linear EST 06-SEP-2000
DEFINITION UI-M-BH2.3-aoh-f-07-0-UI.r1 NIH_BMAP_M_S3.3 Mus musculus cDNA clone UI-M-BH2.3-aoh-f-07-0-UI 5', mRNA sequence.
ACCESSION BE649209
VERSION BE649209.1 GI:9975033
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 421)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution:

Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements The following repetitive elements were found in this cDNA sequence: 356-388, >GC_rich#Low_complexity Seq primer: M13 Reverse.

```
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH2.3-aoh-f-07-0-UI"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NIH_BMAP_M_S3.3"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S3.3 library is a subtracted library of a
                     series, ultimately derived from a mixture of individually
                     tagged normalized libraries from ten regions of the mouse
                     brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hipoccampus) after a series of
                     subtractions to reduce the representation of cDNAs from
                     which ESTs had already been generated. The following
                     serially subtracted libraries were generated in this
                     process: NIH_BMAP_M_S3.3, NIH_BMAP_M_S2, NIH_BMAP_M_S1.
                     The subtracted library (NIH_BMAP_M_S3.3) was constructed
                     as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2
                     clones from which 3' ESTs had been derived was used as a
                     driver in a hybridization with the NIH_BMAP_M_S2 library
                     in the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S3.3 library. This procedure has been
                     previously described (Bonaldo, Lennon and Soares, Genome
                     Research 6:791-806, 1996)"
ORIGIN
```

```
  Query Match             51.6%;  Score 125.8;  DB 10;  Length 421;
  Best Local Similarity   85.9%;  Pred. No. 7e-22;
  Matches  152;  Conservative    0;  Mismatches   22;  Indels    3;  Gaps    1;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||||||||||||||||||||||| |||||||||||||| || | |||||| |||||
Db        343 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACGGCCGCAGCCGGTAAACT 284

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              || ||||| |||||||||||||||| | |||| ||  ||||| | | | ||||||
Db        283 TTCGCAGGCGGCTCAGCCATTGGCTGGTAGAGTCAAGT---CCTCCCCTCGAGCCCTCCG 227

Qy        121 CCCTCCCGCCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCC 177
              ||  |||||||| ||||||||||||||||||||||||||||||||||||||| || ||
Db        226 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCC 170
```

```
RESULT 2
AW121149
LOCUS       AW121149                 343 bp    mRNA    linear   EST 22-OCT-1999
DEFINITION  UI-M-BH2.3-aoh-f-07-0-UI.s1 NIH_BMAP_M_S3.3 Mus musculus cDNA clone
            UI-M-BH2.3-aoh-f-07-0-UI 3', mRNA sequence.
ACCESSION   AW121149
VERSION     AW121149.1  GI:6096482
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            The following repetitive elements were found in this cDNA sequence:
            13-50, >GC_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH2.3-aoh-f-07-0-UI"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NIH_BMAP_M_S3.3"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S3.3 library is a subtracted library of a
                     series, ultimately derived from a mixture of individually
                     tagged normalized libraries from ten regions of the mouse
                     brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hipoccampus) after a series of
                     subtractions to reduce the representation of cDNAs from
                     which ESTs had already been generated. The following
                     serially subtracted libraries were generated in this
                     process: NIH_BMAP_M_S3.3, NIH_BMAP_M_S2, NIH_BMAP_M_S1.
                     The subtracted library (NIH_BMAP_M_S3.3) was constructed
                     as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2
                     clones from which 3' ESTs had been derived was used as a
                     driver in a hybridization with the NIH_BMAP_M_S2 library
                     in the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S3.3 library. This procedure has been
                     previously described (Bonaldo, Lennon and Soares, Genome
                     Research 6:791-806, 1996)
                     TAG_TISSUE=brain-stems
                     TAG_LIB=NIH_BMAP_M_S3.3
                     TAG_SEQ=TCATG"
ORIGIN
```

```
  Query Match             50.9%;  Score 124.2;  DB 10;  Length 343;
  Best Local Similarity   85.3%;  Pred. No. 1.7e-21;
  Matches  151;  Conservative    0;  Mismatches   23;  Indels    3;  Gaps    1;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGCCGGCAAACT 60
              |||||||||||||||||||||||| |||||||||||||| || | |||||| |||||
Db         66 AGCGCCGGGGATTTAAATGCCACTGAAACGGTGATCCATCACGGCCGCAGCCGGTAAACT 125

Qy         61 TTTGCAGGAGGCTCAGCCATTGGCTGACACCGTCACGTGCCCCTCCTCCAGCGTCCTCCG 120
              || ||| | |||||||||||||||| | |||| ||  ||||| | | | ||||||
Db        126 TTCGCATGCGGCTCAGCCATTGGCTGGTAGAGTCAAGT---CCTCCCCTCGAGCCCTCCG 182
```

```
Qy        121 CCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCC 177
              ||  |||||||| ||||||||||||||||||||||||||||||||||||||| || ||
Db        183 CCTGCCCGCCCCGCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCTGGTCC 239


RESULT 3
BM838150
LOCUS       BM838150                625 bp    mRNA    linear   EST 06-MAR-2002
DEFINITION  K-EST0114344 S9SNU601 Homo sapiens cDNA clone S9SNU601-73-G08 5',
            mRNA sequence.
ACCESSION   BM838150
VERSION     BM838150.1  GI:19194559
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 73  row: G  column: 08
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S9SNU601-73-G08"
                     /sex="M"
                     /tissue_type="Ascites"
                     /cell_type="Epithelial"
                     /cell_line="SNU-601"
                     /lab_host="Top10F'"
                     /clone_lib="S9SNU601"
                     /note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI;
                     Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including SfiI
                     site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized with Superscript II using SfiI
                     oligo-dT primer. After first strand synthesis, RNA was
                     degraded by NaOH treatment and cDNA was amplified by PCR
                     reaction. The PCR products were digested with SfiI and
                     cloned into DraIII- digested pME18S-FL3 vector. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
ORIGIN

  Query Match           36.1%;  Score 88;  DB 12;  Length 625;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-12;
  Matches   88;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        157 ATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGC 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGC 60

Qy        217 TTCGCAGGACCAGGTCACTCAGTGACAG 244
              ||||||||||||||||||||||||||||
Db         61 TTCGCAGGACCAGGTCACTCAGTGACAG 88


RESULT 4
AK083575
LOCUS       AK083575                2475 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 9 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:D030047C07 product:homeo box A1, full
            insert sequence.
ACCESSION   AK083575
VERSION     AK083575.1  GI:26350638
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2475)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
```

```
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES              Location/Qualifiers
     source           1. .2475
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="FANTOM_DB:D030047C07"
                      /db_xref="MGI:2418936"
                      /db_xref="taxon:10090"
                      /clone="D030047C07"
                      /tissue_type="whole body"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="9 days embryo"
     CDS              98. .1108
                      /note="unnamed protein product; homeo box A1
                      (MGD|MGI:96170, GB|NM_010449, evidence: BLASTN, 99%,
                      match=2216)
                      putative"
                      /codon_start=1
                      /protein_id="BAC38956.1"
                      /db_xref="GI:26350639"
                      /translation="MDNARMNSFLEYPILGSGDSGTCSARAYPSDHGITTFQSCAVSA
                      NSCGGDDRFLVGRGVQISSPHHHHHHHHHHHPQTATYQTSGNLGISYSHSSCGPSYGA
                      QNFSAPYGPYGLNQEADVSGGYPPCAPAVYSGNLSTPMVQHHHHHQGYAGGTVGSPQY
                      IHHSYGQEQQTLALATYNNSLSPLHASHQEACRSPASETSSPAQTFDWMKVKRNPPKT
                      GKVGEYGYVGQPNAVRTNFNTKQLTELEKEFHFNKYLTRARRVEIAASLQLNETQVKI
                      WFQNRRMKQKKREKEGLLPISPATPPGSDEKTEESSEKSSPSPSAPSPASSTSDTLTT
                      SH"
ORIGIN

  Query Match             32.8%;  Score 80;  DB 11;  Length 2475;
  Best Local Similarity   89.6%;  Pred. No. 6.6e-10;
  Matches   86;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        149 ACATTCATATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCA 208
              ||||||||||||||||||||||| || || |||||||||||||||||||||||||| ||||
Db          2 ACATTCATATCATTTTTCTTCTCTGGTCCTATGGAGGAAGTGAGAAAGTTGGCACGGTCA 61

Qy        209 CGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244
              | | |||||||||||||| |   ||||||||||||||
Db         62 CCCAGGGCTTCGCAGGATCCAATCACTCAGTGACAG 97


RESULT 5
AA199907
LOCUS       AA199907                 428 bp    mRNA    linear   EST 23-JAN-1997
DEFINITION  zq52a07.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
            clone IMAGE:645204 5' similar to TR:G1195540 G1195540 HOXA1
            (ALTERNATIVELY SPLICED) . ;, mRNA sequence.
ACCESSION   AA199907
VERSION     AA199907.1  GI:1795641
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 195.
FEATURES              Location/Qualifiers
     source           1. .428
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:5215455"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:645204"
                      /dev_stage="Ntera-2/RA neuroepithelial cells"
                      /lab_host="SOLR (kanamycin resistant)"
                      /clone_lib="Stratagene neuroepithelium (#937231)"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; Cloned unidirectionally.  Primer: Oligo dT. NT2
                      cells (Ntera-2/c1.D1) induced with Retinoic Acid for 24
                      hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

  Query Match             31.6%;  Score 77;  DB 9;  Length 428;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-09;
  Matches   77;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        168 TCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACC 227
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACC 60

Qy        228 AGGTCACTCAGTGACAG 244
              |||||||||||||||||
Db         61 AGGTCACTCAGTGACAG 77


RESULT 6
BB655265
LOCUS       BB655265                 604 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB655265 RIKEN full-length enriched, 9 days embryo Mus musculus
            cDNA clone D030047C07 5', mRNA sequence.
ACCESSION   BB655265
VERSION     BB655265.1  GI:16489093
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
```

Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .604
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="D030047C07"
/dev_stage="9 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 9 days embryo"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 31.3%; Score 76.4; DB 10; Length 604;
Best Local Similarity 88.3%; Pred. No. 3.9e-09;
Matches 83; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      151 ATTCATATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACG 210
            |||||||||||||||||||| || || ||||||||||||||||||||||||| |||||
Db        1 ATTCATATCATTTTTCTTCTCTGGTCCTATGGAGGAAGTGAGAAAGTTGGCACGGTCACC 60

Qy      211 CCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244
            | |||||||||| || |   ||||||||||||||
Db       61 CAGGGCTTCGCATGATCCAATCACTCAGTGACAG 94
```

RESULT 7
BB846631
LOCUS BB846631 432 bp mRNA linear EST 26-NOV-2001
DEFINITION BB846631 RIKEN full-length enriched, adult male kidney Mus musculus cDNA clone F530003M03 5', mRNA sequence.
ACCESSION BB846631
VERSION BB846631.1 GI:17085006
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 432)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .432
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="F530003M03"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male kidney"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. "

ORIGIN

```
  Query Match             29.9%;  Score 73;  DB 10;  Length 432;
  Best Local Similarity   88.8%;  Pred. No. 2.7e-08;
  Matches   79;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        156 TATCATTTTTCTTCTCCGGCCCCATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGG 215
              ||||||||||||||| || || |||||||||||||||||||||||| ||||| | |||
Db          3 TATCATTTTTCTTCTCTGGTCCTATGGAGGAAGTGAGAAAGTTGGCACGGTCACCCAGGG 62

Qy        216 CTTCGCAGGACCAGGTCACTCAGTGACAG 244
              |||||||||| |   |||||||||||||
Db         63 CTTCGCAGGATCCAATCACTCAGTGACAG 91


RESULT 8
BM462913
LOCUS       BM462913                1104 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6447587 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5537563
            5', mRNA sequence.
ACCESSION   BM462913
VERSION     BM462913.1  GI:18511953
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1104)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12229  row: a  column: 20
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .1104
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5537563"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
ORIGIN

  Query Match             25.8%;  Score 63;  DB 12;  Length 1104;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-05;
  Matches   63;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        182 GAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGA 241
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 GAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGA 70

Qy        242 CAG 244
              |||
Db         71 CAG 73


RESULT 9
AU127876
LOCUS       AU127876                 851 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU127876 NT2RP2 Homo sapiens cDNA clone NT2RP2002264 5', mRNA
            sequence.
ACCESSION   AU127876
VERSION     AU127876.1  GI:10988230
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 851)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .851
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2002264"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RP2"
                     /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                     cells after 2-weeks retinoic acid (RA) induction"
ORIGIN

  Query Match             25.4%;  Score 62;  DB 9;  Length 851;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-05;
  Matches   62;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        183 AGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGAC 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGAC 60

Qy        243 AG 244
              ||
Db         61 AG 62


RESULT 10
AU132057
LOCUS       AU132057                 773 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU132057 NT2RP3 Homo sapiens cDNA clone NT2RP3003711 5', mRNA
            sequence.
ACCESSION   AU132057
VERSION     AU132057.1  GI:10992411
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
```

```
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES              Location/Qualifiers
     source           1. .773
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="NT2RP3003711"
                      /cell_type="teratocarcinoma"
                      /cell_line="NT2"
                      /clone_lib="NT2RP3"
                      /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                      cells after 2-weeks retinoic acid (RA) induction"
ORIGIN

  Query Match             22.5%;  Score 55;  DB 9;  Length 773;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        190 GAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 55


RESULT 11
AA070261
LOCUS       AA070261                471 bp    mRNA    linear   EST 23-DEC-1997
DEFINITION  zm69c05.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
            clone IMAGE:530888 5' similar to SW:HXA1_MOUSE P09022 HOMEOBOX
            PROTEIN HOX-A1 ;, mRNA sequence.
ACCESSION   AA070261
VERSION     AA070261.1  GI:1577621
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.

            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 440.
FEATURES              Location/Qualifiers
     source           1. .471
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:3920400"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:530888"
                      /dev_stage="Ntera-2/RA neuroepithelial cells"
                      /lab_host="SOLR (kanamycin resistant)"
                      /clone_lib="Stratagene neuroepithelium (#937231)"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; Cloned unidirectionally.  Primer: Oligo dT. NT2
                      cells (Ntera-2/c1.D1) induced with Retinoic Acid for 24
                      hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

  Query Match             22.1%;  Score 54;  DB 9;  Length 471;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        191 AGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAAAGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 54


RESULT 12
AU128573
LOCUS       AU128573                640 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU128573 NT2RP2 Homo sapiens cDNA clone NT2RP2003721 5', mRNA
            sequence.
ACCESSION   AU128573
VERSION     AU128573.1  GI:10988927
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES              Location/Qualifiers
     source           1. .640
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="NT2RP2003721"
                      /cell_type="teratocarcinoma"
                      /cell_line="NT2"
                      /clone_lib="NT2RP2"
                      /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                      cells after 2-weeks retinoic acid (RA) induction"
ORIGIN

  Query Match             20.5%;  Score 50;  DB 9;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 0.024;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        195 AGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTGGCACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 50
```

```
RESULT 13
CA415315
LOCUS       CA415315                661 bp    mRNA    linear   EST 07-NOV-2002
DEFINITION  UI-H-EZ0-baw-j-13-0-UI.s1 NCI_CGAP_Ch1 Homo sapiens cDNA clone
            UI-H-EZ0-baw-j-13-0-UI 3', mRNA sequence.
ACCESSION   CA415315
VERSION     CA415315.1  GI:24777966
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of
            Orthoapedics
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 1-43, >POLY_A#Simple_repeat (matched compliment) 413-491,
            >GC_rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-EZ0-baw-j-13-0-UI"
                     /tissue_type="Chondrosarcoma Grade II"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Ch1"
                     /note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_Ch1 is a cDNA library containing the following
                     tissue(s): Chondrosarcoma Grade II. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     TGATCACGCT.
                     TAG_TISSUE=grade-2-chondrosarcoma
                     TAG_LIB=UI-H-EZ0
                     TAG_SEQ=ATCTAATATG"
ORIGIN

  Query Match             18.3%;  Score 44.6;  DB 14;  Length 661;
  Best Local Similarity   92.2%;  Pred. No. 0.6;
  Matches   47;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACTGCGGAAGC 51
              |||||||||||||||||||||||||||||||||||||||||| | | |||
Db        610 AGCGCCGGGGATTTAAATGCCACTAAAACGGTGATCCATCACCGTTGGAGC 660


RESULT 14
AL650594
LOCUS       AL650594                678 bp    mRNA    linear   EST 07-NOV-2003
DEFINITION  AL650594 XGC-neurula Silurana tropicalis cDNA clone TNeu028f22 5',
            mRNA sequence.
ACCESSION   AL650594
VERSION     AL650594.2  GI:38223211
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
  TITLE     Sanger Xenopus tropicalis EST project 2001 (11_2003)
  JOURNAL   Unpublished (2003)
COMMENT     On Dec 13, 2001 this sequence version replaced gi:17660265.
            Contact: Huckle E
            Sanger Institute
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
            cDNA was oligo dT primed from 5ug of poly A+ RNA from neurula.
            EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
            5' end and NotI at the 3' end.
            Vector: pCS107; Site_1: EcoRI; Site_2: NotI
            Host: Escherichia coli DH10B
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TNeu028f22.p1kSP6
            Sequencing primer: SP6.
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TNeu028f22"
                     /dev_stage="neurula"
                     /lab_host="Escherichia coli DH10B"
                     /clone_lib="XGC-neurula"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from neurula.
                     EcoRI-NotI cut cDNA was then ligated into pCS107 with
                     EcoRI at the 5' end and NotI at the 3' end."
ORIGIN

  Query Match             18.0%;  Score 44;  DB 9;  Length 678;
  Best Local Similarity   72.0%;  Pred. No. 0.86;
  Matches   72;  Conservative    0;  Mismatches   25;  Indels    3;  Gaps    1;

Qy        118 CCGCCCTCCCGCCCCCCCTCTTGCGCACTGTACATTCATATCATTTTTCTTCTCCGGCCC 177
              | | |||| |  |     |    | ||||||||||||||||| ||||   |   | || ||
Db        103 CTGGCCTCTCCACGTGACCTGCCCTCACTGTACATTCATATGATTT---TGTGCTGGACC 159

Qy        178 CATGGAGGAAGTGAGAAAGTTGGCACAGTCACGCCGGGCT 217
              |||||||||||||||||||||||| | |||||| ||||
Db        160 TATGGAGGAAGTGAGAAAGTTGGCACGGCCACGCCAGGCT 199


RESULT 15
AU129261
LOCUS       AU129261                550 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU129261 NT2RP2 Homo sapiens cDNA clone NT2RP2005190 5', mRNA
            sequence.
ACCESSION   AU129261
VERSION     AU129261.1  GI:10989615
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
```

```
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2005190"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RP2"
                     /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                     cells after 2-weeks retinoic acid (RA) induction"
ORIGIN

  Query Match             17.6%;  Score 43;  DB 9;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        202 ACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244
              |||||||||||||||||||||||||||||||||||||||||||
Db          1 ACAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 43


Search completed: September 18, 2004, 13:16:52
Job time : 2065.12 secs
```